

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:01 ; Search time 3102.12 seconds
(without alignments)
3130.089 Million cell updates/sec

Title: US-09-617-174b-1_copy_451_914

Perfect score: 464

Sequence: 1 tgaataatcctaatacacagg.....gcaacttcgcgcgaatcat 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pi:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match Length	ID	Description

1	462.4	99.7	170269	9	AC090307	AC090307 Homo sapi
2	462.4	99.7	173131	2	AC036176	AC036176 Homo sapi
3	462.4	99.7	173836	2	AP001404	AP001404 Homo sapi
4	390	84.1	214212	2	AC015793	AC015793 Homo sapi
5	340.8	73.4	12426	6	AX251236	AX251236 Sequence
6	313.6	67.6	12426	6	AX251235	AX251235 Sequence
7	75.2	16.2	214212	2	AC015793	AC015793 Homo sapi
8	44.4	9.6	170136	9	AC003082	AC003082 Human BAC
9	44.4	9.6	225053	2	AC084396	AC084396 Homo sapi
10	42.6	9.2	164339	2	AC025792	AC025792 Homo sapi
11	42.4	9.1	166982	9	AC084783	AC084783 Homo sapi
12	42.2	9.1	22601	8	AB022222	AB022222 Arabidops
13	41.4	8.9	7218	6	I66494	I66494 Sequence 14
14	41.4	8.9	195470	2	AC022259	AC022259 Homo sapi
15	41.2	8.9	122258	2	AL356427	AL356427 Homo sapi
16	40.6	8.8	154235	9	HS428A13	282209 Human DNA s
17	40.6	8.8	159906	2	AL590706	AL590706 Homo sapi
18	40.6	8.8	160517	2	AL589785	AL589785 Homo sapi
19	40.6	8.8	183704	2	AC023326	AC023326 Homo sapi
20	40.4	8.7	108253	9	AL136164	AL136164 Human DNA
21	40.2	8.7	89747	9	AL591022	AL591022 Human DNA
22	40	8.6	131711	2	AL445194	AL445194 Homo sapi
23	40	8.6	194515	2	AC095027	AC095027 Canis fam
24	39.8	8.6	113880	3	PFMAL3P4	AL008970 Plasmodiu
25	39.8	8.6	166125	9	AC023932	AC023932 Homo sapi
26	39.6	8.5	10127	10	AF250369S6	AF249893 Mus muscu
27	39.6	8.5	98905	2	AC095617	AC095617 Rattus no
28	39.6	8.5	110000	2	AC098517	AC098517 Rattus no
29	39.6	8.5	133751	9	HS1131C10	AL121931 Human DNA
30	39.4	8.5	74992	2	AC069574	AC069574 Homo sapi
31	39.4	8.5	154235	9	HS428A13	282209 Human DNA s
32	39.4	8.5	160517	2	AL589785	AL589785 Homo sapi
33	39.4	8.5	183704	9	AC023326	AC023326 Homo sapi
34	39.4	8.5	185676	9	AC073927	AC073927 Homo sapi
35	39.2	8.4	24023	9	AC105392	AC105392 Homo sapi
36	39.2	8.4	153003	2	AC094798	AC094798 Rattus no
37	39.2	8.4	172051	30	AC044914	AC044914 Homo sapi
38	39	8.4	26203	2	PFMAL13P7	AL096784 Plasmodiu
39	39	8.4	43547	3	U28971	U28971 Caenorhabdi
40	39	8.4	160596	2	AC107082	AC107082 Homo sapi
41	39	8.4	165046	2	AC091991	AC091991 Homo sapi
42	39	8.4	169755	2	AP001199	AP001199 Homo sapi
43	39	8.4	172915	2	AC099742	AC099742 Papio cyn
44	39	8.4	188207	2	AC021800	AC021800 Homo sapi
45	38.8	8.4	213793	9	AC019212	AC019212 Homo sapi

ALIGNMENTS

RESULT	1	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
LOCUS	AC090307	Homo sapiens	chromosome 18, clone RP11-851B10, complete sequence.			
DEFINITION	AC090307	Homo sapiens	chromosome 18, clone RP11-851B10, complete sequence.			
ACCESSION	AC090307	HTG.				
VERSION	AC090307.7	GI:16041402				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 170269)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.					
TITLE	Homo sapiens chromosome 18, clone RP11-851B10					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 170269)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barin, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dekrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,					

repeat_region	/rpt_family="AluSx" 3218. .3365
repeat_region	/rpt_family="MIR" 3565. .3590
repeat_region	/rpt_family="(TG)n" complement(3782. .4028)
repeat_region	/rpt_family="MIR" 4036. .4177
repeat_region	/rpt_family="MER5A" 4601. .5620
repeat_region	/rpt_family="HSMAR2" 5621. .5643
repeat_region	/rpt_family="(CAA)n" 5644. .5866
repeat_region	/rpt_family="HSMAR2" 6184. .6475
repeat_region	/rpt_family="L1MA2" 6677. .6837
repeat_region	/rpt_family="(TA)n" 6852. .7018
repeat_region	/rpt_family="(TA)n" 6994. .7122
repeat_region	/rpt_family="(CATATA)n" 7146. .7332
repeat_region	/rpt_family="(CATATA)n" 7371. .7526
repeat_region	/rpt_family="(TA)n" complement(7527. .7835)
repeat_region	/rpt_family="AluSx" 7936. .8080
repeat_region	/rpt_family="MIR" 8697. .8927
repeat_region	/rpt_family="AluSx" complement(9036. .9288)
repeat_region	/rpt_family="AluSg" 9873. .9934
repeat_region	/rpt_family="(GA)n" 10340. .10367
repeat_region	/rpt_family="AT_rich" 10368. .10686
repeat_region	/rpt_family="AluY" 10687. .10711
repeat_region	/rpt_family="AT_rich" complement(10870. .11043)
repeat_region	/rpt_family="L2" 11348. .11375
repeat_region	/rpt_family="(T)n" 12359. .12449
repeat_region	/rpt_family="L3" complement(12634. .12890)
repeat_region	/rpt_family="MIR" complement(13436. .13725)
repeat_region	/rpt_family="AluJb" 14244. .14297
repeat_region	/rpt_family="(CGGGG)n" complement(15188. .15350)
repeat_region	/rpt_family="MIR3" 16056. .16114
repeat_region	/rpt_family="MIR3" complement(16129. .16429)
repeat_region	/rpt_family="AluSx" 16532. .16646
repeat_region	/rpt_family="MER94" 16855. .17145
repeat_region	/rpt_family="L2" 17377. .17572
repeat_region	/rpt_family="MIR3" complement(20036. .20330)
repeat_region	/rpt_family="Charliel" complement(20333. .20486)
repeat_region	/rpt_family="Charliel" 20625. .20691
repeat_region	/rpt_family="L2" complement(20691. .20844)

FEATURES	source	Location/Qualifiers
repeat_region	1. .170269	
repeat_region		/organism="Homo sapiens"
repeat_region		/db_xref="taxon:9606"
repeat_region		/chromosome="18"
repeat_region		/map="18"
repeat_region		/clone="RP11-851B10"
repeat_region		/clone_lib="RPC1-11 Human Male BAC"
repeat_region		complement(2..177)
repeat_region		/rpt_family="L1MC5"
repeat_region		complement(189..489)
repeat_region		/rpt_family="Alusx"
repeat_region		complement(708..803)
repeat_region		/rpt_family="MIR"
repeat_region		complement(1616..1904)
repeat_region		/rpt_family="AluJo"
repeat_region		complement(2815..3120)

```

repeat_region complement(20829..21075)
repeat_region /rpt_family="MLTIC"
repeat_region complement(21104..21336)
repeat_region /rpt_family="MLTIC"
repeat_region complement(22112..22391)
repeat_region /rpt_family="AluX"
repeat_region 22506..22581
repeat_region /rpt_family="Tigres5"
repeat_region complement(23483..23664)
repeat_region /rpt_family="MIR"
repeat_region 23979..24190
repeat_region /rpt_family="MER20"
repeat_region complement(25404..25455)
repeat_region /rpt_family="MER5B"
repeat_region complement(25769..25853)
repeat_region /rpt_family="AluS"
unsure 25964..25998
unsure /note="Single clone coverage"
unsure 25999..26004
unsure /note="<30 qual single clone coverage"
unsure 26005..26009
unsure /note="Single clone coverage"
repeat_region complement(26204..26297)
repeat_region /rpt_family="MERSA"
repeat_region 27103..27222
repeat_region /rpt_family="MIR3"
repeat_region 27281..27573
repeat_region /rpt_family="AluY"
repeat_region 28033..28160
repeat_region /rpt_family="MIR"
repeat_region complement(28205..28490)
repeat_region /rpt_family="L1PB2"
repeat_region 28491..28801

Query Match 99.7% Score 462.4; DB 9; Length 170269;
Best Local Similarity 99.8%; Pred. No. 2:1e-11;
Matches 463; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tgaagtaactcattcacaggatttataaaagaacttcctgcgcaccattacaacaata 60
DB 13726 TGAGTAATCTAATCACAGGATTTTAAAGAAGAACTTCTGCGCCACCATTAAACAATA 13785

QY 61 tctctaccacatttggtagtaataatttttgcctaagtagtacctaatttttagtaggcact 120
DB 13786 TCTCCTACCAATTTGGTAGTAATAATTTTGGCTAATAGTACCTAATTTTAGTAGGCAC 13845

QY 121 gttttatatacatatccattcctcttttttgattgtctttctgtttaattgagcagcta 180
DB 13846 GTGTTTATACATATATCCATTCTCTTTTGTGATTGCTTTCTGTTTAAATGGCAGCTA 13905

QY 181 cctctctggcattcagagaatgagctgctgagtttacacaaaagaatgagatcag 240
DB 13906 CTTCTCTTGGCATCTAGCAGAAATGAGTCTGCAGTTTACACAAAAGAAATGGAGATCAG 13965

QY 241 agtactttttgcccacaaacgtgtctgagaaatttgtagttagttactatcacacacatt 300
DB 13966 AGTACTTTTGGCCACCACCAACCTGCTGAGAAATTTGAGTCTACTATCATCACACATT 14025

QY 301 acttttatttcataaataatttcacacttcctgcgtcgtgcgtgcgtgcgtgcgtgcgt 360
DB 14026 ACTTTTATTTTCATGAATAATTTTACGCTTTCGCTGCGCGGAGAGGATTCGCGTA 14085

QY 361 ccatgctctgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 420
DB 14086 CCGATGCTGTCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 14145

QY 421 cttttggaagctgtgcagacaacagcaacttcagcctgaatcat 464
DB 14146 CTTTGGGAAGCTGTCAGACAACAGTAACCTTCAGCCTGAATCAT 14189

RESULT 2
AC036176/c

```

LOCUS AC036176 173131 bp DNA linear HTG 13-DEC-2001

DEFINITION Homo sapiens chromosome 18 clone RP11-635N19 map 18, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.

AC036176

AC036176.7 GI:17647019

KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVIFEIN.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 173131)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE Young Submission

JOURNAL

COMMENT Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 13, 2001 this sequence version replaced gi:16949388.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L9215

Center clone name: 635_N_19

* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 114304: contig of 114304 bp in length

* 114305 114404: gap of 100 bp

* 114403 173131: contig of 58727 bp in length.

Location/Qualifiers

1..173131

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-635N19"

/clone_lib="RPCI-11 Human Male BAC"

51584 a 35316 c 35660 g 50470 t 101 others

BASE COUNT


```
RESULT 5
AX251236/C
LOCUS AX251236 12426 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 204 from Patent WO0168912.
ACCESSION AX251236
VERSION AX251236.1 GI:15984659
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 12426)
Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with tumor suppressor genes and
TITLE oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 73.4%; Score 340.8; DB 6; Length 12426;
Best Local Similarity 83.4%; Pred. No. 7.2e-81;
Matches 387; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 1 tgagtaactcctaatcacagagattttaaaagaacctctcgccaccctaaacaata 60
Db 8056 TAAATATCTATCAATAAATTTTAAAAAATACTTCTACGCCACCCATTAAACAATA 7997
QY 61 tctcctaccatttgtagtaataattttgctaagtagtaacctaattttaggcact 120
Db 7996 TCTCTTACCCTTAAATAAATAATTTTACTAATAATACCTAAATTTTAAATAAACACT 7937
QY 121 gttgttatacatatccattctctctttttgtagtctcttctgttttaaggcgacta 180
Db 7936 ATATTATACATATATCCATCTCTCTTTTAAATTTATCTTCTATTATTAACAACATA 7877
QY 181 cctcctctggcactcgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 7876 CCTCTCTTAACATCTAACAAAAATAACTACTACAAATTTTACAAAAAATAAATAACAA 7817
QY 241 agtactttttgtgcaccacgctgctgagaaatttgtagttgtagttactatcatcacact 300
Db 7816 AATACCTTTTATACCAACCAACGATCTCAAAAAATTTATAATATTACTATCATCACACAT 7757
QY 301 acttttatttcacgaattttcaccttccggtccggtcggtggccgagagattgcccgt 360
Db 7756 ACTTTTATTCATCGAATATTTTACCTTCGATCCTACGTAACCCGAAAAAATTACCGTA 7697
QY 361 cgcattgtctgtagtgcgtgaactcacagcccttctcgcgcgaacatgttggagggc 420
Db 7696 CGCATATCTATCGTATACATATAAATCAACACCCCTTCTACCCGCAATATTAATAAAC 7637
QY 421 cttttggaagctgtgcagacaacagcaacttcaagctgaatcat 464
Db 7636 CTTTAAAAAACTATACAAACAACAATAACTTCAACCTAAATCAT 7593
RESULT 6
AX251235
LOCUS AX251235 12426 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 203 from Patent WO0168912.
ACCESSION AX251235
VERSION AX251235.1 GI:15984658
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 12426)
Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with tumor suppressor genes and
TITLE oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 73.4%; Score 340.8; DB 6; Length 12426;
Best Local Similarity 83.4%; Pred. No. 7.2e-81;
Matches 387; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 1 tgagtaactcctaatcacagagattttaaaagaacctctcgccaccctaaacaata 60
Db 8056 TAAATATCTATCAATAAATTTTAAAAAATACTTCTACGCCACCCATTAAACAATA 7997
QY 61 tctcctaccatttgtagtaataattttgctaagtagtaacctaattttaggcact 120
Db 7996 TCTCTTACCCTTAAATAAATAATTTTACTAATAATACCTAAATTTTAAATAAACACT 7937
QY 121 gttgttatacatatccattctctctttttgtagtctcttctgttttaaggcgacta 180
Db 7936 ATATTATACATATATCCATCTCTCTTTTAAATTTATCTTCTATTATTAACAACATA 7877
QY 181 cctcctctggcactcgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 7876 CCTCTCTTAACATCTAACAAAAATAACTACTACAAATTTTACAAAAAATAAATAACAA 7817
QY 241 agtactttttgtgcaccacgctgctgagaaatttgtagttgtagttactatcatcacact 300
Db 7816 AATACCTTTTATACCAACCAACGATCTCAAAAAATTTATAATATTACTATCATCACACAT 7757
QY 301 acttttatttcacgaattttcaccttccggtccggtcggtggccgagagattgcccgt 360
Db 7756 ACTTTTATTCATCGAATATTTTACCTTCGATCCTACGTAACCCGAAAAAATTACCGTA 7697
QY 361 cgcattgtctgtagtgcgtgaactcacagcccttctcgcgcgaacatgttggagggc 420
Db 7696 CGCATATCTATCGTATACATATAAATCAACACCCCTTCTACCCGCAATATTAATAAAC 7637
QY 421 cttttggaagctgtgcagacaacagcaacttcaagctgaatcat 464
Db 7636 CTTTAAAAAACTATACAAACAACAATAACTTCAACCTAAATCAT 7593
```

```
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 203 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3527 a 137 c 2591 g 6171 t
ORIGIN
Query Match 57.6%; Score 313.6; DB 6; Length 12426;
Best Local Similarity 79.7%; Pred. No. 1.4e-73;
Matches 370; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 1 tgagtaactcctaatcacagagattttaaaagaacctctcgccaccctaaacaata 60
Db 4371 TGAGTAATTTTAAATATAGGATTTTAAAAAGAAATTTTTCGCTTATTTTAAATAATA 4430
QY 61 tctcctaccatttgtagtaataattttgctaagtagtaacctaattttaggcact 120
Db 4431 TTTTATTATTAATTTGCTAGTAAATATTTTCTTAATAGTATTTTAAATTTTAGGTAGTATT 4490
QY 121 gttgttatacatatccattctctctttttgtagtctcttctgttttaaggcgacta 180
Db 4491 GGTGTATATATATATTTATTTTTCATTTGATTTGTTTTTTCGTTATTTTATGAGGTAGTTA 4550
QY 181 cctcctctggcactcgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 4551 TTTTTCGCTTATTTAGTAGAATGAGCTTCTGTAGTTTATATATAAAGAAATGAGATTAG 4610
QY 241 agtactttttgtgcaccacgctgctgagaaatttgtagttactatcatcacact 300
Db 4611 AGTATTTTTCGTTGTTTATTAACCGTGTTCGAAAAATTTTGTAGTGTATTTATATATATAT 4670
QY 301 acttttatttcacgaattttcaccttccggtccggtcggtggccgagagattgcccgt 360
Db 4671 ATTTTTATTTTATCGAATATTTTATTTTTCGGTTTCGGTGGTGGCGAGAGGATTGTCGTA 4730
QY 361 cgcattgtctgtagtgcgtgaactcacagcccttctcgcgcgaacatgttggagggc 420
Db 4731 CGTATCTTTGTCAGTATGATGTAATTTATAGTTTTTTTTTGTCCGAATATGTTGGAGGT 4790
QY 421 cttttggaagctgtgcagacaacagcaacttcaagctgaatcat 464
Db 4791 TTTTTCGAAAGTTGCTGATAGATAATAGTAATTTTATGTTTGAATAT 4834
RESULT 7
AX251793
LOCUS AC015793 214212 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-1117D15
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
```

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meidrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 695: contig of 695 bp in length
696 795: gap of 100 bp
796 1513: contig of 718 bp in length
1514 1613: gap of 100 bp
1614 2337: contig of 724 bp in length
2338 2437: gap of 100 bp
2438 3161: contig of 724 bp in length
3162 3261: gap of 100 bp
3262 3966: contig of 705 bp in length
3967 4066: gap of 100 bp
4067 4772: contig of 706 bp in length
4773 4872: gap of 100 bp
4873 5564: contig of 692 bp in length
5565 5664: gap of 100 bp
5665 6397: contig of 733 bp in length
6398 6497: gap of 100 bp
6498 7202: contig of 705 bp in length
7203 7302: gap of 100 bp
7303 8028: contig of 726 bp in length
8029 8128: gap of 100 bp
8129 8803: contig of 675 bp in length
8804 8903: gap of 100 bp
8904 9613: contig of 710 bp in length
9614 9713: gap of 100 bp
9714 10409: contig of 695 bp in length
10410 10509: gap of 100 bp
10510 11237: contig of 728 bp in length
11238 11337: gap of 100 bp
11338 12043: contig of 706 bp in length
12044 12143: gap of 100 bp
12144 12867: contig of 724 bp in length
12868 12967: gap of 100 bp
12968 13660: contig of 693 bp in length
13661 13760: gap of 100 bp
13761 14489: contig of 729 bp in length
14490 14589: gap of 100 bp
14590 15314: contig of 725 bp in length

15315 15414: gap of 100 bp
15415 16129: contig of 715 bp in length
16130 16229: gap of 100 bp
16230 16946: contig of 717 bp in length
16947 17046: gap of 100 bp
17047 17769: contig of 723 bp in length
17770 17869: gap of 100 bp
17870 18574: contig of 705 bp in length
18575 18674: gap of 100 bp
18675 19379: contig of 705 bp in length
19380 19479: gap of 100 bp
19480 20186: contig of 707 bp in length
20187 20286: gap of 100 bp
20287 20980: contig of 694 bp in length
20981 21080: gap of 100 bp
21081 21779: contig of 699 bp in length
21780 21879: gap of 100 bp
21880 22594: contig of 715 bp in length
22595 22694: gap of 100 bp
22695 23399: contig of 705 bp in length
23400 23499: gap of 100 bp
23500 24217: contig of 718 bp in length
24218 24317: gap of 100 bp
24318 25026: contig of 709 bp in length
25027 25126: gap of 100 bp
25127 25836: contig of 710 bp in length
25837 25936: gap of 100 bp
25937 26637: contig of 701 bp in length
26638 26737: gap of 100 bp
26738 27446: contig of 709 bp in length
27447 27546: gap of 100 bp
27547 28274: contig of 728 bp in length
28275 28374: gap of 100 bp
28375 29095: contig of 721 bp in length
29096 29195: gap of 100 bp
29196 29917: contig of 722 bp in length
29918 30017: gap of 100 bp
30018 30724: contig of 707 bp in length
30725 30824: gap of 100 bp
30825 31529: contig of 705 bp in length
31530 31629: gap of 100 bp
31630 32324: contig of 695 bp in length
32325 32424: gap of 100 bp
32425 33116: contig of 692 bp in length
33117 33216: gap of 100 bp
33217 33936: contig of 720 bp in length
33937 34036: gap of 100 bp
34037 34750: contig of 714 bp in length
34751 34850: gap of 100 bp
34851 35578: contig of 728 bp in length
35579 35678: gap of 100 bp
35679 36393: contig of 715 bp in length
36394 36493: gap of 100 bp
36494 37203: contig of 710 bp in length
37204 37303: gap of 100 bp
37304 37965: contig of 662 bp in length
37966 38065: gap of 100 bp
38066 38781: contig of 716 bp in length
38782 38881: gap of 100 bp
38882 39590: contig of 709 bp in length
39591 39690: gap of 100 bp
39691 40403: contig of 713 bp in length
40404 40503: gap of 100 bp
40504 41238: contig of 735 bp in length
41239 41338: gap of 100 bp
41339 42064: contig of 726 bp in length
42065 42164: gap of 100 bp
42165 42880: contig of 716 bp in length
42881 42980: gap of 100 bp
42981 43687: contig of 707 bp in length
43688 43787: gap of 100 bp
43788 44504: contig of 717 bp in length
44505 44604: gap of 100 bp

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 225053)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 225053)
Waterston,R.H.
Direct Submission
Submitted (30-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 5, 2001 this sequence version replaced gi:12863223.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0134N03
----- Summary Statistics -----
Sequencing vector: M13; 91%
Sequencing vector: plasmid; 4%
Chemistry: Dye-terminator Big Dye; 4% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184917 bases at least Q40
Consensus quality: 196549 bases at least Q30
Consensus quality: 203318 bases at least Q20
Insert size: 196000; agarose-ff
Insert size: 219053; sum-of-contigs
Quality coverage: 3.56 in Q20 bases; agarose-ff
Quality coverage: 2.91 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4207: contig of 4207 bp in length
* 4208 4307: gap of unknown length
* 4308 10132: contig of 5825 bp in length
* 10133 10232: gap of unknown length
* 10233 15820: contig of 5588 bp in length
* 15821 15920: gap of unknown length
* 15921 20334: contig of 4414 bp in length
* 20335 20434: gap of unknown length
* 20435 27845: contig of 7411 bp in length
* 27846 32177: contig of 4232 bp in length
* 32178 32277: gap of unknown length
* 32278 38504: contig of 6227 bp in length
* 38505 38604: gap of unknown length
* 38605 47769: contig of 9165 bp in length
* 47770 47869: gap of unknown length
* 47870 55641: contig of 7772 bp in length
* 55642 55741: gap of unknown length
* 55742 65119: contig of 9378 bp in length
* 65120 65219: gap of unknown length
* 65220 73704: contig of 8485 bp in length
* 73705 73804: gap of unknown length
* 73805 83677: contig of 9873 bp in length
* 83678 83777: gap of unknown length
* 83778 92194: contig of 8417 bp in length
* 92195 92294: gap of unknown length
* 92295 100513: contig of 8219 bp in length
* 100514 106583: contig of 5970 bp in length
* 106584 106683: gap of unknown length
* 106684 117190: contig of 10507 bp in length
* 117191 117290: gap of unknown length
* 117291 117291: contig of 1438 bp in length

118729 118828: gap of unknown length
118829 120265: contig of 1437 bp in length
120266 120365: gap of unknown length
120366 121652: contig of 1287 bp in length
121653 121752: gap of unknown length
121753 122860: contig of 1108 bp in length
122861 122960: gap of unknown length
122961 124298: contig of 1338 bp in length
124299 124398: gap of unknown length
124399 125617: contig of 1219 bp in length
125618 125717: gap of unknown length
125718 127326: contig of 1609 bp in length
127327 127426: gap of unknown length
127427 129643: contig of 2217 bp in length
129644 129743: gap of unknown length
129744 130956: contig of 1213 bp in length
130957 131057: gap of unknown length
131057 132284: contig of 1228 bp in length
132285 132384: gap of unknown length
132385 134025: contig of 1641 bp in length
134026 134125: gap of unknown length
134126 135771: contig of 1646 bp in length
135772 135871: gap of unknown length
135872 137168: contig of 1297 bp in length
137169 137268: gap of unknown length
137269 138776: contig of 1508 bp in length
138777 138876: gap of unknown length
138877 140630: contig of 1754 bp in length
140631 140730: gap of unknown length
140731 143390: contig of 2660 bp in length
143391 143490: gap of unknown length
143491 144885: contig of 1395 bp in length
144886 144985: gap of unknown length
144986 146277: contig of 1292 bp in length
146278 146377: gap of unknown length
146378 148790: contig of 2413 bp in length
148791 148890: gap of unknown length
148891 151618: contig of 2728 bp in length
151619 151718: gap of unknown length
151719 153528: contig of 1810 bp in length
153529 153628: gap of unknown length
153629 155632: contig of 2004 bp in length
155633 155732: gap of unknown length
155733 158353: contig of 2621 bp in length
158354 158453: gap of unknown length
158454 161521: contig of 3068 bp in length
161522 161821: gap of unknown length
161822 163929: contig of 2308 bp in length
163930 164029: gap of unknown length
164030 165640: contig of 1611 bp in length
165641 165740: gap of unknown length
165741 168599: contig of 2959 bp in length
168600 168799: gap of unknown length
168800 169971: contig of 1172 bp in length
169972 170071: gap of unknown length
170072 171383: contig of 1312 bp in length
171384 171483: gap of unknown length
171484 174613: contig of 3130 bp in length
174614 174713: gap of unknown length
174714 178023: contig of 3310 bp in length
178024 178123: gap of unknown length
178124 179809: contig of 1686 bp in length
179810 179909: gap of unknown length
179910 182166: contig of 2257 bp in length
182167 182266: gap of unknown length
182267 184778: contig of 2512 bp in length
184779 184878: gap of unknown length
184880 187289: contig of 2411 bp in length
187290 187389: gap of unknown length
187390 190025: contig of 2636 bp in length
190026 190125: gap of unknown length
190126 193628: contig of 3503 bp in length
193629 193728: gap of unknown length

```

* 193729 196385: contig of 2657 bp in length
* 196386 196485: gap of unknown length
* 196486 199287: contig of 2802 bp in length
* 199288 199387: gap of unknown length
* 199388 202073: contig of 2686 bp in length
* 202074 202173: gap of unknown length
* 202174 205921: contig of 3748 bp in length
* 205922 206021: gap of unknown length
* 206022 211281: contig of 5260 bp in length
* 211282 211381: gap of unknown length
* 211382 215785: contig of 4404 bp in length
* 215786 215885: gap of unknown length
* 215886 220968: contig of 5083 bp in length
* 220969 221068: gap of unknown length
* 221069 225053: contig of 3985 bp in length.
FEATURES
    source
        1. .225053
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosomes="UNK"
            /clone="RP11-134N3"
            /note="assembly_name:Contig100"
            /clone_end:77
            vector_side:left
            4308. .10132
            /note="assembly_name:Contig101"
            10233. .15820
            /note="assembly_name:Contig102"
            15921. .20334
            /note="assembly_name:Contig103"
            20435. .27845
            /note="assembly_name:Contig104"
            27946. .32177
            /note="assembly_name:Contig105"
            32278. .38504
            /note="assembly_name:Contig106"
            38605. .47769
            /note="assembly_name:Contig107"
            47870. .55641
            /note="assembly_name:Contig108"
            55742. .65119
            /note="assembly_name:Contig109"
            65220. .73704

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```

Query Match          9.6%; Score 44.4; DB 2; Length 225053;
Best Local Similarity 46.7%; Pred. No. 0.26;
Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

```

QY 1 t9agtaatcctaatacagagattttaaagaactctgcgcacccattaaacaata 60
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194153 TAAGTAATCTCAGTACATATTTCTAGAAAAAGAGATACAAATATTTTCTCTTGC 194094
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 tctctaccaatttggtagtaataattttgctaatagtacctaatttttagtgtaggcact 120
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194093 ACTCTATAGTGCTTCTCAAAATATTTTTTAAAGTCATCTTAGTTTTTCTTATA 194034
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 gtgtttatacatatccattctcttttttgattgcttcttctgttaatgggcagcta 180
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194033 TTCACAATGCTTAATAAGATTTTGTAGACTCTCATGCTGAATGATGTGTAGATT 193974
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 cctctctggcacttagcagaatgagctgctgcagtttacacaaaagaatggagatcag 240
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193973 CAACTGTGCTAGTTTAAACACACAGCTATGCTCATCTCTTTGTGAAGCTTAATATATGTG 193914
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 agtactttttgtccaccaagctgtctgagaaatttgtagtcttactatcatcacatt 300
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193913 TCTGTCTTTTATTCCAATTAGTCCCTGTGTAAGTCTTGTGTACAAATATTCACAGTTT 193854
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 301 ac 302
Db 193853 CC 193852

```

RESULT 10

AC025792/c

LOCUS

DEFINITION

AC025792

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC025792 164339 bp DNA Linear HTG 06-APR-2000
Homo sapiens chromosome 11 clone RP11-243M7 map 11, WORKING DRAFT
SEQUENCE, 13 unordered pieces.
AC025792
HTG: HTGS_PHASE1; HTGS_DRAFT.
AC025792.2 GI:7467488
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164339)
Homo sapiens chromosome 11, clone RP11-243M7
Unpublished

2 (bases 1 to 164339)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhgaltier, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collins, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
McDrum, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Plerre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Yong, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 6, 2000 this sequence version replaced gi:7239678.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L7901
Center clone name: 243_M7

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 156382 bases at least Q40
Consensus quality: 160561 bases at least Q30
Consensus quality: 162162 bases at least Q20

Insert size: 170000; agarose-fp
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
consists of 13 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will


```

DEFINITION Homo sapiens chromosome 2 clone RP11-335E2 map 2, *** SEQUENCING IN
AC022259
AC022259.1 GI:6778490
VERSION
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195470)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-335E2
Unpublished
2 (bases 1 to 195470)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F.,
Boquslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cooke, P.,
DeAtrellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5217
Center clone name: 335_E_2
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1169: contig of 1169 bp in length
* gap of unknown length
* 1170 2744: contig of 1575 bp in length
* gap of unknown length
* 2745 3891: contig of 1147 bp in length
* gap of unknown length
* 3892 4958: contig of 1067 bp in length
* gap of unknown length
* 4959 6042: contig of 1084 bp in length
* gap of unknown length
* 6043 7112: contig of 1070 bp in length
* gap of unknown length
* 7113 8242: contig of 1130 bp in length
* gap of unknown length
* 8243 9440: contig of 1198 bp in length
* gap of unknown length
* 9441 10503: contig of 1063 bp in length
* gap of unknown length
* 10504 11766: contig of 1263 bp in length
*
* 11767 gap of unknown length
* 12798: contig of 1032 bp in length
* gap of unknown length
* 14299: contig of 1501 bp in length
* gap of unknown length
* 15314: contig of 1015 bp in length
* gap of unknown length
* 16978: contig of 1664 bp in length
* gap of unknown length
* 18049: contig of 1071 bp in length
* gap of unknown length
* 19597: contig of 1548 bp in length
* gap of unknown length
* 21003: contig of 1406 bp in length
* gap of unknown length
* 22126: contig of 1123 bp in length
* gap of unknown length
* 23502: contig of 1376 bp in length
* gap of unknown length
* 24450: contig of 948 bp in length
* gap of unknown length
* 25766: contig of 1316 bp in length
* gap of unknown length
* 27235: contig of 1469 bp in length
* gap of unknown length
* 29226: contig of 1991 bp in length
* gap of unknown length
* 30961: contig of 1735 bp in length
* gap of unknown length
* 32608: contig of 1647 bp in length
* gap of unknown length
* 34331: contig of 1723 bp in length
* gap of unknown length
* 36533: contig of 2202 bp in length
* gap of unknown length
* 39299: contig of 2766 bp in length
* gap of unknown length
* 41451: contig of 2152 bp in length
* gap of unknown length
* 43295: contig of 1844 bp in length
* gap of unknown length
* 45678: contig of 2383 bp in length
* gap of unknown length
* 48551: contig of 2873 bp in length
* gap of unknown length
* 51186: contig of 2635 bp in length
* gap of unknown length
* 53207: contig of 2021 bp in length
* gap of unknown length
* 55697: contig of 2490 bp in length
* gap of unknown length
* 58328: contig of 2631 bp in length
* gap of unknown length
* 59424: contig of 1096 bp in length
* gap of unknown length
* 62209: contig of 2785 bp in length
* gap of unknown length
* 64795: contig of 2586 bp in length
* gap of unknown length
* 67972: contig of 3177 bp in length
* gap of unknown length
* 71707: contig of 3735 bp in length
* gap of unknown length
* 74928: contig of 3221 bp in length
* gap of unknown length
* 78667: contig of 3739 bp in length
* gap of unknown length
* 82959: contig of 4292 bp in length
* gap of unknown length
* 88848: contig of 5889 bp in length
* gap of unknown length
* 93807: contig of 4959 bp in length
* gap of unknown length

```



```
source
1. .122258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_lib="RP11-78B14"
/clone_lib="RPC1-11.1"
1. .2843
/note="assembly_fragment:00918
fragment_chain:1"
2944. .6529
/note="assembly_fragment:00793
fragment_chain:1"
6630. .16752
/note="assembly_fragment:00592
fragment_chain:1"
16853. .19610
/note="assembly_fragment:00690
fragment_chain:1"
19711. .30027
/note="assembly_fragment:00317
fragment_chain:2"
30128. .35982
/note="assembly_fragment:00020
fragment_chain:2"
36083. .39034
/note="assembly_fragment:00521
fragment_chain:2"
39135. .41364
/note="assembly_fragment:00105
fragment_chain:3"
41465. .45036
/note="assembly_fragment:00723
fragment_chain:3"
45137. .54043
/note="assembly_fragment:00499
fragment_chain:4"
54144. .67915
/note="assembly_fragment:00122
fragment_chain:4"
68016. .71519
/note="assembly_fragment:00673
fragment_chain:5"
71620. .81674
/note="assembly_fragment:00183
fragment_chain:5"
81775. .85041
/note="assembly_fragment:00848
fragment_chain:6"
85142. .88867
/note="assembly_fragment:00370
fragment_chain:6"
88968. .94857
/note="assembly_fragment:00286"
94958. .98108
/note="assembly_fragment:00287"
98209. .100727
/note="assembly_fragment:00352"
100828. .104494
/note="assembly_fragment:00571"
104595. .107100
/note="assembly_fragment:00818"
107201. .110041
/note="assembly_fragment:00860"
110142. .114180
/note="assembly_fragment:00929"
114281. .116760
/note="assembly_fragment:00932"
116861. .122258
/note="assembly_fragment:01078"
37169 a 23367 c 22456 g 36922 t 2344 others
BASE COUNT
ORIGIN
```

```
Query Match      8.9%; Score 41.2; DB 2; Length 122258;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 61 tctctaccaattggtagtaataatttggctaataatgtaacctaatttttaggtaggcact 120
Db 4356 TCACCTAACAGTTGGCTATTTTAAATTTATTCATTTTAAAACTTTGAAGATTTTCTG 4297

Qy 121 gtgtttatacatatccattccctcttttttgattgtctcttctgtttaaattgggcagcta 180
Db 4296 GAAATGATATGCTTCGTGCTAGTCTCTTCATTTCTTCTCTGTTTAACTGCCATGGC 4237

Qy 181 cctctcttggcatotagcagaatgagctgctgcagttttacacaaaaaagaatggagatcag 240
Db 4236 ACAGAAATTTTCAGATATTGGATCTAACAAAAAAGATATGCTATAATAACAAACAGGTAAA 4177

Qy 241 agtactttttgtgccaccaacgctgctgagaaatttggtagttactatcatcacacatt 300
Db 4176 ATACATTTTGGGATTACATTAATAAAAAAGAAACTATAGGCATTTCCAGACTCTGAAGTT 4117

Qy 301 actttttatttcac 314
Db 4116 ATGATAATATTTTC 4103
```

Search completed: September 24, 2002, 15:52:52
Job time: 20267 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:46:36 ; Search time 403.26 Seconds
(without alignments)
1975.520 Million cell updates/sec

Title: US-09-617-174B-1-COPY_451_914

Perfect score: 464

Sequence: 1 tgaataatcctaatcacagg.....gcaacttcagcctgaatcat 464

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	464	100.0	1141	18	AA*89000 Human maspin promo
2	340.8	73.4	12426	22	AA*46482 Tumour suppressor
3	313.6	67.6	12426	22	AA*46481 Tumour suppressor
4	38.2	8.2	1239	22	AA*68963 Human immune/haema
5	38.2	8.2	13537	22	AA*68964 Human immune/haema
6	38	8.2	56153	22	AA*46793 Tumour suppressor
7	37.4	8.1	6063	22	AA*46337 Tumour suppressor
8	37	8.0	7143	21	AA*70250 Plasmodium falcipar
9	36.8	7.9	513445	22	AA*161373 Soybean 318013 reg

10	36.6	7.9	8964	24	ABL333534 Human immune syste
11	36.6	7.9	8964	24	AA*61276 Human immune regulat
c 12	36.4	7.8	44211	22	AA*85974 Human immune/haema
c 13	36.4	7.8	61020	22	AA*46787 Tumour suppressor
14	36.2	7.8	2234	17	AA*09022 Arabidopsis thalia
15	36.2	7.8	6309	24	ABL32304 Human immune syste
16	35.8	7.7	6127	24	ABL33615 Human immune syste
17	35.8	7.7	8123	24	ABL34227 Human immune syste
c 18	35.6	7.7	952	19	AA*14118 H. pylori GHPO 122
c 19	35.6	7.7	9293	22	AA*53320 Chemically pretrea
20	35.4	7.6	7403	22	AA*46804 Tumour suppressor
21	35.4	7.6	7403	22	ABL34217 Human immune syste
22	35.4	7.6	56153	22	AA*46794 Chemically pretrea
23	35.2	7.6	11047	22	AA*45480 Human immune syste
24	35.2	7.6	11047	24	ABL33985 Human immune/haema
25	35.2	7.6	28091	22	AA*15375 Human nervous syst
26	35.2	7.6	28091	22	AA*77432 Human immune/haema
27	35.2	7.6	28120	22	AA*15379 Human immune/haema
28	35.2	7.6	28120	22	AA*77434 Human genome from
29	35.2	7.6	114793	22	AA*08215 Tumour suppressor
30	35	7.5	15592	22	AA*46454 Human immune syste
31	35	7.5	15592	24	ABL33327 Human immune syste
c 32	35	7.5	16509	24	ABL33321 Enterococcus faeca
c 33	34.8	7.5	1494	20	AA*20170 Human immune syste
c 34	34.8	7.5	6134	24	ABL33183 Enterococcus faeca
c 35	34.8	7.5	6259	20	AA*13066 Human immune syste
c 36	34.8	7.5	7069	22	AA*46654 Tumour suppressor
37	34.8	7.5	7069	24	ABL33353 Human immune syste
38	34.8	7.5	7069	24	AA*61220 Human gene regulat
c 39	34.6	7.5	370	22	AA*184576 Human polynucleoti
c 40	34.6	7.5	890	22	AA*15375 Human breast cance
c 41	34.6	7.5	6136	24	AA*61297 Human gene regulat
42	34.6	7.5	18028	23	ABL13284 Drosophila melanog
43	34.4	7.4	834	22	AA*03576 Human cDNA clone (
44	34.4	7.4	5947	22	AA*46676 Tumour suppressor
c 45	34.4	7.4	6127	24	ABL33614 Human immune syste

ALIGNMENTS

RESULT 1

AA*89000

ID AA*89000 standard; cDNA; 1141 BP.

XX AC AA*89000;

XX DT 22-APR-1998 (first entry)

XX DE Human maspin promoter and partial coding cDNA.

XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

XX KW tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT promoter 1..1141

FT FT /tag= a

FT FT /note= "maspin promoter region"

FT FT misc_signal 443..449

FT FT /tag= b

FT FT /note= "Putative regulatory element AP2"

FT FT misc_signal 451..457

FT FT /tag= c

FT FT /note= "Putative regulatory element API"

FT FT misc_signal 495..490

FT FT /tag= d

FT FT /note= "Putative regulatory element Ets"

FT FT misc_signal 684..691

FT FT /tag= e

FT FT /note= "Putative regulatory element HRE"

FT FT misc_signal 846..851

RESULT 4
ID AAK68963 standard; DNA; 1239 BP.
XX AC AAK68963;
XX DT 06-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23775.
XX DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01354.
XX PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184684.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

```
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure: SEQ ID NO 23775; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 1239 BP; 391 A; 191 C; 176 G; 481 T; 0 other;

Query Match      8.2%; Score 38.2; DB 22; Length 1239;
Best Local Similarity 48.4%; Pred. NO. 0.61;
Matches 106; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 1  tgaagtaatcctaatcacaggattttaaagaacattcttcgcgccaccattaaacaata 60
Db 82  ttagttttatcatcattgaagacatttgaagatattctttattcaagcttaaatcccat 141

Qy 61  tctctaccatttggtagtaaatatttgcctaagtagtaccttaatttttagtaggcact 120
Db 142  ttcttaattaaattagacaacattctttacaagaaattgaataatctgtagtactg 201

Qy 121  ggtttatcatatcatcattcttcttttttgattgtcttcttcttaattgagcagcta 180
Db 202  ttcatagaatacaaatgttagcctttatttggtttttttaaggaattggcagctg 261

Qy 181  cctctcttggcattcagcagaatgagctgctcagttta 219
Db 262  ttgtgaggtcataagaagacgagtaagagagtaattta 300

RESULT      5
AAK68964
ID  AAK68964 standard; DNA; 13537 BP.
XX
XX AAK68964;
```

```
XX
XX 06-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:23776.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0236279.
XX 22-AUG-2000; 2000US-0236681.
XX 22-AUG-2000; 2000US-0236868.
XX 22-AUG-2000; 2000US-0237182.
XX 23-AUG-2000; 2000US-0237009.
XX 30-AUG-2000; 2000US-0238924.
XX 01-SEP-2000; 2000US-0239287.
XX 01-SEP-2000; 2000US-0239343.
XX 01-SEP-2000; 2000US-0239344.
XX 05-SEP-2000; 2000US-0239509.
XX 05-SEP-2000; 2000US-0239513.
XX 06-SEP-2000; 2000US-0239513.
XX 06-SEP-2000; 2000US-0239437.
XX 08-SEP-2000; 2000US-0239438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
```

```
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2000US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX Disclosure; SEQ ID NO 23776; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX Sequence 13537 BP; 4245 A; 2216 C; 2328 G; 4748 T; 0 other;

Query Match 8.2%; Score 38.2; DB 22; Length 13537;
Best Local Similarity 48.4%; Pred. NO. 1.4; Indels 0; Gaps 0;
Matches 106; Conservative 0; Mismatches 113;

QY 1 tgaagtaactcctacagagattttaaagaacactctctgcgcaccattaaacaata 60
Db 3161 tagttttatcatcattgaagacattgttaagtatatcttttattcaaaagcttaaatccat 3220
QY 61 tctcctaccaatttgtagtaaaattttgctaaatagtaacctaatttttagtaggcact 120
Db 3221 ttcttaattaaattagacaacactttcttacaagaaaattgaataatctgtagtagct 3280
QY 121 gtgtttatcatatccattctctcttttttgattgtcttctgttttaattgggcagctta 180
Db 3281 ttcatagaatcaaatgtagccctttattttgttttttttttaagcattggcagctg 3340
QY 181 cctctctggcatctagcagaatgagctgtgcagttta 219
Db 3341 ttgtgaagtcataagaacgcgagtaagaggtaattta 3379

RESULT 6
AAS46793
ID AAS46793 standard; DNA; 56153 BP.
XX AAS46793;
AC AAS46793;
XX 18-DEC-2001 (first entry)
DT Tumour suppressor gene derived chemically modified sequence #519.
DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW
```

KW cytosine methylation; ds.
 XX Homo sapiens.
 OS
 PN WO200168912-A2.
 XX
 XX 20-SEP-2001.
 PD
 XX
 XX 15-MAR-2001; 2001WO-EP02955.
 PF
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR
 PR 06-APR-2000; 2000DE-1019058.
 PR
 PR 07-APR-2000; 2000DE-1019173.
 PR
 PR 30-JUN-2000; 2000DE-1032529.
 PR
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602752/68.
 DR
 XX
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 PT
 XX
 XX Claim 1; SEQ ID No 519; 27pp; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 56153 BP; 13943 A; 1002 C; 14095 G; 27113 T; 0 other;
 SQ
 Query Match 8.2%; Score 38; DB 22; Length 56153;
 Best Local Similarity 47.5%; Pred. No. 2.8;
 Matches 113; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
 QY 73 ttggttagtaaatatttgcctaataagcacttaatttttaggtaggcactgtgttatacat 132
 Db 31726 ttgggattatcaggatgagttattgttttttagttttgttttttttttttttttttt 31785
 QY 133 atatccattcctcttt 192
 Db 31786 ttattgagtt 31845
 QY 193 tctagcagaatagctgctgcagttttacacaaaagaatggagatcagagttactttttgt 252
 Db 31846 ttctcgtagtataatt 31905
 QY 253 gccaccacagctgctgagaaattgttagtttactatcatcacacattacttttttttt 310

Db 31906 ttaggtt 31963
 RESULT 7
 AAS46337
 ID AAS46337 standard; DNA; 6063 BP.
 XX
 XX AAS46337;
 AC
 XX 18-DEC-2001 (first entry)
 DT
 XX Tumour suppressor gene derived chemically modified sequence #59.
 DE
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 KW
 XX Homo sapiens.
 OS
 XX WO200168912-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 15-MAR-2001; 2001WO-EP02955.
 PF
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR
 PR 06-APR-2000; 2000DE-1019058.
 PR
 PR 07-APR-2000; 2000DE-1019173.
 PR
 PR 30-JUN-2000; 2000DE-1032529.
 PR
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602752/68.
 DR
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 PT
 XX
 XX Claim 1; SEQ ID No 59; 27pp; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 6063 BP; 1704 A; 41 C; 1173 G; 3145 T; 0 other;
 SQ
 Query Match 8.1%; Score 37.4; DB 22; Length 6063;

PT cytosine methylation -

PT disease -

XX PS Disclosure; SEQ ID No 237; 26pp; English.

CC The invention relates to 224 nucleic acid sequences comprising at least

CC 18 bases of a chemically pretreated gene associated with gene regulation

CC selected from 43 known genes (or complementary sequences). The

CC chemical pretreatment converts cytosine bases unmethylated at the

CC 5-position to uracil or another base with hybridisation behaviour

CC dissimilar to cytosine, to enable analysis of cytosine methylations.

CC The DNA sequences, oligomers (or sets/arrays) and method are

CC useful in the diagnosis of diseases (or predisposition to diseases)

CC associated with gene regulation and in therapy of such diseases, by

CC enabling analysis of the cytosine methylation patterns of such genes,

CC kits are provided. They are especially useful in diagnosis

CC and therapy of e.g. severe combined immunodeficiency disease, cardiac

CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,

CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,

CC preclampsia, graft versus-host disease. The present sequence is a

CC sequence included in the sequence data for this specification and is

CC associated with the human gene regulation-associated genes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 8964 BP; 2596 A; 89 C; 1815 G; 4464 T; 0 other;

Query Match 7.9%; Score 36.6; DB 24; Length 8964;

Best Local Similarity 53.1%; Pred. No. 3.6;

Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 72 ttgtgtagtaaatatttgcctaagtacctaatttttagtgtaggcactgtgtttatca 131

Db 7681 ttaattagtaaatatttgcctaagtacctaatttttagtgtaggcactgtgtttatca 131

QY 132 tatatccattccctctttttttgattgtcttttctttaaaggcagctacctcttggc 191

Db 7741 ttttttggtagattgtttgtatttttttaggggtttttttttttttttttttat 7800

QY 192 atctagcagaatgagctgcagttt 218

Db 7801 ttgtaggttgtagtcggttttagttt 7827

RESULT 12

AAK85974/C

ID AAK85974 standard; DNA; 44211 BP.

XX AC AAK85974;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40786.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225271.

PR 14-AUG-2000; 2000US-0225275.

PR 14-AUG-2000; 2000US-0225275.

PR 14-AUG-2000; 2000US-0225275.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:10 ; Search time 3309.28 Seconds
(without alignments)
1892.432 Million cell updates/sec

Title: US-09-617-174B-1-COPY_451_914

Perfect score: 464

Sequence: 1 tsggtaacccaatcacagg.....gcaacttcagcggaaatcat 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estopl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.4	9.4	892	A2539346	A2539346 ENTGP83TR
2	41.8	9.0	1201	CNS016FX	AL106695 Drosophil
3	40.6	8.8	681	A0421326	AQ421326 RPCI-II-1
4	40.2	8.7	1042	CNS024NB	AL180992 Tetraodon
5	40	8.6	1006	CNS03ESG	AL240793 Tetraodon
6	39.8	8.6	774	AL546807	AL546807 AL546807
7	39.6	8.5	457	BE943660	BE943660 UI-M-BH3-
8	39.6	8.5	494	A1848569	A1848569 UI-M-AMI-
9	39.6	8.5	1142	BM464476	BM464476 AGENCOURT
10	39.2	8.4	928	CNS00DKY	AL071865 Drosophil
11	39	8.4	850	A2673135	A2673135 ENTLI30TR
12	38.8	8.4	648	BH587629	BH587629 BOHEX88TF
13	38.6	8.3	365	A1605224	A1605224 ve78a05_x
14	38.6	8.3	388	B58764	B58764 CIT-RSP-201
15	38.6	8.3	419	BM163067	BM163067 EST565590
16	38.6	8.3	711	BF468105	BF468105 UI-M-CD0-
17	38.4	8.3	443	AQ450690	AQ450690 HS_5161_A

c	18	38.4	8.3	507	12	CNS01852	AL108896 Drosophil
	19	38.4	8.3	726	10	BM170969	BM170969 EST573492
c	20	38.4	8.3	884	12	CNS00CXG	AL060063 Drosophil
c	21	38.4	8.3	1099	12	CNS01481	AL103836 Drosophil
	22	38.4	8.3	1101	12	CNS0039G	AL063921 Drosophil
	23	38.2	8.2	641	12	AG066397	AG066397 Pan trogl
	24	38.2	8.2	907	12	CNS07COV	AL439445 T7 end of
	25	37.8	8.1	535	12	A2748844	A2748844 RPCI-24-1
c	26	37.8	8.1	878	12	CNS0187R	AL108993 Drosophil
c	27	37.8	8.1	952	12	CNS0180Y	AL108748 Drosophil
c	28	37.8	8.1	985	12	CNS00J34	AL076343 Drosophil
c	29	37.8	8.1	1023	12	CNS00CVF	AL059889 Drosophil
c	30	37.8	8.1	1101	12	CNS008X3	AL052544 Drosophil
c	31	37.6	8.1	312	12	A2821437	A2821437 2M0094009
c	32	37.6	8.1	396	12	AQ683613	AQ683613 HS_5458_B
c	33	37.6	8.1	873	12	CNS024M5	AL180950 Tetraodon
c	34	37.4	8.1	797	10	BF253656	BF253656 HVSMEF000
c	35	37.4	8.1	1007	12	CNS01413	AL104181 Drosophil
c	36	37.4	8.1	1101	12	CNS0006J	AL062049 Drosophil
c	37	37.2	8.0	353	9	AW627276	AW627276 pa28a06.y
c	38	37.2	8.0	569	10	BE886609	BE886609 601509355
c	39	37.2	8.0	607	12	AG068424	AG068424 Pan trogl
c	40	37.2	8.0	614	9	AL655660	AL655660 AL655660
c	41	37.2	8.0	761	12	BH470789	BH470789 BOGYM20TR
c	42	37.2	8.0	767	12	CNS00AQX	AL055924 Drosophil
c	43	37.2	8.0	865	12	A2675018	A2675018 ENTHS57TR
c	44	37.2	8.0	889	12	A2543817	A2543817 ENTFT53TF
c	45	37	8.0	500	9	AU087167	AU087167 AU087167

ALIGNMENTS

RESULT 1
A2539346
LOCUS ENTGP83TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
DEFINITION genomic, DNA sequence.
ACCESSION A2539346
VERSION A2539346.1 GI:11145135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 892)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 847.
Location/Qualifiers
1..892
/organism="Entamoeba histolytica".
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

[illegible][illegible]

RESULT 3
 AQ421326
 LOCUS
 DEFINITION
 'DNA sequence.
 accession
 version
 keywords
 source
 organism

REFERENCE AUTHORS	1 (bases 1 to 681) Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL COMMENT	Unpublished (1997) Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html . Seq primer: SP6

```

FEATURES
source
Location/Qualifiers
l..68l
/organism="Homo sapiens"
/db_xref="GB:75896"
/db_xref="taxon:1606"
/clone="RPC11-10710"
/clone_lib="RPC1-11"
/sex="Male"
/tumortissues=""

```

BASE COUNT ORIGIN	149 a	98 c	107 g	326 t	1 others
Query Match	8.8%; Score 40.6; DB 12; Length 681;				
Best Local Similarity	46.9%; Pred. No. 25;				
Matches 127; Conservative	0; Mismatches 144; Indels 0; Gaps 0;				
Qy	66	taccaatttggtagtaaatatttgcctaataagtagtaccctaatttttaggtaggcactgtgtt	125		
Db	185	TTCCAATTATTGGCACAAGTGTGATAGTACGACCAATAAGATTCTGTGAATTTCTG	244		
Qy	126	tatacatatataccattcccttcttttttggattgtcttcttcttctttaaaggcagcactcctt	185		
Db	245	CATTATCAGCTGTAATGTATCTTTTTCATTTCTGATTTTATTACTGTATCTCTCTTTT	304		
Qy	186	cttgccatctagcagaatgagctgcgcagtttacacaaaagaatggagatcacagatc	245		
Db	305	ATTCTTAGTCTAAGGTTTGTCAATTTTGTATTAAATTAATAAATTTTATTTCAGTGATA	364		
Qy	246	tttttggccacaacgtgctgcgagaaattttagtcttactatcatcacacattacttt	305		
Db	365	TTTTTATTGTTTCTTCATTTTCAGATTATGATTTATGTTATGTTGCTGATCATACTA	424		
Qy	306	tatttcacgaatatttcaccttccogtctc	336		
Db	425	TGTCCTCTACTATTATTGAGTTTGACTTTCT	455		
RESULT 4					
CNS024NB					
LOCUS					
DEFINITION	CNS024NB 1042 bp DNA linear GSS 12-MAY-2000				
	Tetraodon nigroviridis genome survey sequence T7 end of clone				
	235C14 of library G from Tetraodon nigroviridis, genomic survey				
	sequence.				
ACCESSION	AL180992				
VERSION	AL180992.1 GI:7819049				
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Tetraodon nigroviridis				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;				
	Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 1042)				
AUTHORS	Roest-Crollius H., Jaillon O., Dasilva C., Fizames C., Fisher C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1042)				
AUTHORS	Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissenbach J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 1042)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases				
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .				
FEATURES	Location/Qualifiers				
source	1. .1042				
	/organism="Tetraodon nigroviridis"				
	/db_xref="taxon:99883"				
	/clone="235C14"				
	/clone_lib="G"				
	/note="Genoscope sequence ID : C0AG235BB07LPI-end : T7"				
BASE COUNT	286 a	208 c	256 g	255 t	1 others
ORIGIN					
Query Match	8.6%; Score 40; DB 12; Length 1006;				
Best Local Similarity	65.9%; Pred. No. 34;				
Matches 58; Conservative	0; Mismatches 30; Indels 0; Gaps 0;				
Qy	78	agtaaatttttgctaataagtagcactaatttttaggtaggcagcactgttttatacataatc	137		
Db	386	AATAAATCTATTCTCTAATAATAATAATATTTATTTATATAAATATATATACCTTTAT	445		
BASE COUNT	291 a	224 c	277 g	247 t	3 others
ORIGIN					
Query Match	8.7%; Score 40.2; DB 12; Length 1042;				
Best Local Similarity	52.4%; Pred. No. 30;				
Matches 63; Conservative	0; Mismatches 38; Indels 0; Gaps 0;				
Qy	65	ctaccaatttggtagtaaatatttgcctaataagtagtaccctaatttttaggtaggcactgtgt	124		
Db	310	CTATAAAAAAATAAATAATCTATTCTAATAATAATAATATTTATTTATAAATATTAT	369		
Qy	125	ttatacatatataccattcccttcttttttggattgtcttcttctg	165		
Db	370	ATATACCTTTATTATTTCCTTTTATTTCATTTTCTGCTG	410		
RESULT 5					
CNS03ESG					
LOCUS					
DEFINITION	CNS03ESG 1006 bp DNA linear GSS 17-MAY-2000				
	Tetraodon nigroviridis genome survey sequence T7 end of clone				
	020H04 of library G from Tetraodon nigroviridis, genomic survey				
	sequence.				
ACCESSION	AL240793				
VERSION	AL240793.1 GI:7961562				
KEYWORDS	GSS; genome survey sequence.				
SOURCE	Tetraodon nigroviridis.				
ORGANISM	Tetraodon nigroviridis				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;				
	Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 1006)				
AUTHORS	Roest-Crollius H., Jaillon O., Dasilva C., Fizames C., Fisher C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1006)				
AUTHORS	Roest-Crollius H., Jaillon O., Dasilva C., Bouneau L., Fisher C., Bernot A., Fizames C., Wincker P., Brottier P., Quetier F., Saurin W. and Weissenbach J.				
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 1006)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases				
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .				
FEATURES	Location/Qualifiers				
source	1. .1006				
	/organism="Tetraodon nigroviridis"				
	/db_xref="taxon:99883"				
	/clone="020H04"				
	/clone_lib="G"				
	/note="Genoscope sequence ID : C0BG020DD02LPI-end : T7"				

Query Match	8.5%	Score 39.6	DB 9	Length 494		
Best Local Similarity	56.3%	Pred. No. 44				
Matches	57	Conservative	0	Mismatches 29	Indels 0	Gaps 0
72	ttttagtagaaataattttgctaatagtacctatctttaggttaggcacgtgtttataca	131				
15	TTTGCAAGTCATATTTTACTGATGTAATTAATATTTCTTAGGAATGTTGGTTATCCA	74				
132	tatatccattcctctcttttttgattg	157				
75	ACTACCTATTTTCTTTTACTTTG	100				
RESULT 9						
LOCUS	BM464476/c					
DEFINITION	AGENCOURT_6438833 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535563					
ACCESSION	5', mRNA sequence.					
VERSION	BM464476					
KEYWORDS	EST.					
SOURCE	BM464476.1 GI:18513518					
ORGANISM	human.					
REFERENCE	Homo sapiens					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
JOURNAL	1 (bases 1 to 1142)					
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1223 row: n column: 12 High quality sequence stop: 630. Location/Qualifiers 1. ll42 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5535563" /clone_lib="NIH_MGC_71" /tissue_type="leiomyosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.1 kb." 378 a 245 c 224 g 259 t 36 others					
BASE COUNT						
ORIGIN						
Query Match	8.5%	Score 39.6	DB 10	Length 1142		
Best Local Similarity	48.6%	Pred. No. 41				
Matches	90	Conservative	0	Mismatches 95	Indels 0	Gaps 0
105	tttttaggtaggcactgtgtttatcacatcatcctcttttttgattgctttct	164				
645	TTTTNNNNNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT	586				
165	gttaattggcagctacacctctcttggcatctagcagaatgagctgctgcagttttacaca	224				
585	TTTTTTTGACATATACCACTTACTTATGTACATAAAGGAATGGGAAGGGGAAATG	526				
225	aaagaatggagatcagagtactttttgtgccaccaacgctgtctgagaaattttagtctt	284				
525	AAAGAATACAGAAAACATATACGGTACTAGTCAGGATGTGGTGAACCAAAATTCGAGTTT	466				

AUTHORS Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building

JOURNAL COMMENT Unpublished (1997)
Other_GSSS: CIT-HSP-2013020.TR

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21

Class: BAC ends.

FEATURES

source
1. .388
/organism="Homo sapiens"
/db_xref="GDB:7042534"
/db_xref="taxon:9606"
/clone="2013020"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 137 a 50 c 52 g 149 t
ORIGIN

Query Match 8.3%; Score 38.6; DB 12; Length 388;
Best Local Similarity 52.9%; Pred. No. 76;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 6 aatcctaatacagatttaaaagaacctctgcgccaccattaaacaatactcc 65
Db 314 ACTAATAATACGTGAATTTAAATAAATCTGTAGCAACCAATTCACATATATTA 255
Qy 66 taccaaattgtagtaaatatttgcctaattgtagtaacctaatttttagtaggcactgtgt 125
Db 254 TTATAATATAATATATATAATAATAAAGTCTTATGAGCCATGTGTCTGAATTT 195
Qy 126 tatacatatccatccctctcttttttgattgtcttt 162
Db 194 TTTATAGAAGAACAAATTTTCAATTTAATTCCTTTT 158

RESULT 15
LOCUS BM163067/c 419 bp mRNA linear EST 04-DEC-2001
DEFINITION EST565590 PYBS Plasmodium yoelii yoelii cDNA clone PYCLA48 5' end, mRNA sequence.

ACCESSION BM163067.1 GI:17308748
VERSION BM163067
KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 419)

REFERENCE Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Fraser, C.M. and Carucci, D.J.

Plasmodium yoelii EST project at TIGR
Unpublished (2001)

TITLE Parasite Genomics Group
JOURNAL The Institute for Genomic Research
COMMENT 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

source
Location/Qualifiers
1. .419
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCLA48"
/clone_lib="PYBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybrizAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybrizAP vector and plasmid DNA isolated."

BASE COUNT 226 a 37 c 49 g 107 t
ORIGIN

Query Match 8.3%; Score 38.6; DB 10; Length 419;
Best Local Similarity 51.4%; Pred. No. 76;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 85 atttgctaatagtagtacctaatattttagtaggcactgtgtttatcacatatatccattct 144
Db 277 ATATTATTATAGATAGCCATTTTCTCATTTTAATTTTGTACTTTTAGCTAGCTGTTTC 218
Qy 145 tcttttttgattgtctttctgttttaaggcagctacctctcttgccatctagcagaatg 204
Db 217 ACGTTTTTTTTTTTTTTTTTTTTTTTATTTTCCCTTTACTTCTTTCTGTGTATGATCTTC 158
Qy 205 agctgctgcagtttacacaaaagaatggagatcagagtactttttgtgtccac 257
Db 157 ATTTATTGAAGTTTATTCATACAGAATAATATATCATACATATTTGTGTCTTC 105

Search completed: September 24, 2002, 14:45:15
Job time: 16650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:53:42 : Search time 3102.12 Seconds
(without alignments)
3420.162 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507

Sequence: 1 tggtaatacctaatacacagg.....ccaagaggcttgagtaggag 507

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hgt.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

1	503.8	99.4	170269	9	AC090307	AC090307 Homo sapi
2	503.8	99.4	173131	2	AC036176	AC036176 Homo sapi
3	503.8	99.4	173836	2	AP001404	AP001404 Homo sapi
4	431.4	85.1	214212	2	AX015793	AX015793 Homo sapi
5	362.2	71.4	124236	6	AX251236	AX251236 Sequence
6	343.8	67.8	124236	6	AX251235	AX251235 Sequence
7	116.6	23.0	214212	2	AC015793	AC015793 Homo sapi
8	54.6	10.8	177842	2	AC025047	AC025047 Mus muscu
9	44.4	8.8	170136	9	AC003082	AC003082 Human BAC
10	44.4	8.8	225053	2	AC084396	AC084396 Homo sapi
11	42.6	8.4	164339	2	AC025792	AC025792 Homo sapi
12	42.4	8.4	166982	9	AC084783	AC084783 Homo sapi
13	42.2	8.3	22601	8	AB022222	AB022222 Arabidops
14	41.4	8.2	7218	6	I66494	I66494 Sequence 14
15	41.4	8.2	195470	2	AC022259	AC022259 Homo sapi
16	41.2	8.1	122258	2	AL356427	AL356427 Homo sapi
17	40.6	8.0	154235	9	HS428A13	HS428A13 Human DNA s
18	40.6	8.0	159906	2	AL590706	AL590706 Homo sapi
19	40.6	8.0	160517	2	AL589785	AL589785 Homo sapi
20	40.6	8.0	183704	2	AC023326	AC023326 Homo sapi
21	40.4	8.0	108253	9	AL136164	AL136164 Human DNA
22	40.2	7.9	89747	9	AL591022	AL591022 Human DNA
23	40	7.9	131711	2	AL445194	AL445194 Homo sapi
24	39.8	7.9	194515	2	AC095027	AC095027 Canis fam
25	39.8	7.9	113880	3	PFMAL3P4	PFMAL3P4 Plasmodiu
26	39.6	7.9	166125	9	AC023932	AC023932 Homo sapi
27	39.6	7.8	10127	10	AF250369S6	AF249893 Mus muscu
28	39.6	7.8	98905	2	AC095617	AC095617 Rattus no
29	39.6	7.8	110000	2	AC098517	AC098517 Rattus no
30	39.6	7.8	133751	9	HS1131C10	AL121931 Human DNA
31	39.4	7.8	74992	2	AC069574	AC069574 Homo sapi
32	39.4	7.8	154235	9	HS428A13	HS428A13 Human DNA s
33	39.4	7.8	160517	2	AL589785	AL589785 Homo sapi
34	39.4	7.8	183704	2	AC023326	AC023326 Homo sapi
35	39.4	7.8	185675	9	AC073927	AC073927 Homo sapi
36	39.2	7.7	24023	9	AC105392	AC105392 Homo sapi
37	39.2	7.7	153003	2	AC094798	AC094798 Rattus no
38	39.2	7.7	172051	30	AC044914	AC044914 Homo sapi
39	39	7.7	26203	2	PFMAL13P7	AL096784 Plasmodiu
40	39	7.7	43547	3	U28971	U28971 Caenorhabdi
41	39	7.7	160696	2	AC107082	AC107082 Homo sapi
42	39	7.7	165046	2	AC091991	AC091991 Homo sapi
43	39	7.7	169755	2	AP001199	AP001199 Homo sapi
44	39	7.7	172915	2	AC099742	AC099742 Papilo cyn
45	39	7.7	188207	2	AC021800	AC021800 Homo sapi

ALIGNMENTS

RESULT	1	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
LOCUS	AC090307	Homo sapiens	chromosome 18,	clone RP11-851B10,	complete	sequence.
DEFINITION	AC090307	Homo sapiens	chromosome 18,	clone RP11-851B10,	complete	sequence.
ACCESSION	AC090307	Homo sapiens	chromosome 18,	clone RP11-851B10,	complete	sequence.
VERSION	AC090307.7	GI:16041402				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 170269)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.					
TITLE	Homo sapiens chromosome 18,					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 170269)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boughey,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,					

Jones, C., Karatas, A., Laroque, K., Lamazares, R., Landers, T.,
Lehotzky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnez, C., Suberant, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 170269)

REFERENCE
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepl, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Feerleira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginder, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Katat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 11, 2001 this sequence version replaced gi:15290867.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: L12576
Center clone name: 851_B_10

----- Location/Qualifiers -----
1. .170269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-851B10"
/clone_1lb="RP11-11 Human Male BAC"
/complement(2..177)
/rpt_family="L1MC5"
/complement(189..489)
/rpt_family="AluSx"
/complement(708..803)
/rpt_family="MIR"
/complement(1616..1904)
/rpt_family="AluJo"
/complement(2815..3120)

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

FEATURES
source
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region

/rpt_family="AluSx"
3218..3365
/rpt_family="MIR"
3565..3590
/rpt_family="(TG)n"
complement(3782..4028)
/rpt_family="MIR"
4036..4177
/rpt_family="MERSA"
4601..5620
/rpt_family="HSMAR2"
5621..5643
/rpt_family="(CAA)n"
5644..5866
/rpt_family="HSMAR2"
6184..6475
/rpt_family="L1MA2"
6677..6837
/rpt_family="(TA)n"
6852..7018
/rpt_family="(TA)n"
6994..7122
/rpt_family="(CATATA)n"
7146..7332
/rpt_family="(CATATA)n"
7371..7526
/rpt_family="(TA)n"
complement(7527..7895)
/rpt_family="AluSx"
7936..8080
/rpt_family="MIR"
8697..8927
/rpt_family="AluSx"
complement(9036..9288)
/rpt_family="AluSx"
9873..9934
/rpt_family="(GA)n"
10340..10367
/rpt_family="AT-rich"
10368..10686
/rpt_family="AluY"
10687..10711
/rpt_family="AT-rich"
complement(10870..11043)
/rpt_family="L2"
11348..11375
/rpt_family="(T)n"
12359..12449
/rpt_family="L3"
complement(12634..12890)
/rpt_family="MIR"
complement(13436..13725)
/rpt_family="AluJb"
14244..14297
/rpt_family="(CGGG)n"
complement(15188..15350)
/rpt_family="MIR3"
16056..16114
/rpt_family="MIR3"
complement(16129..16429)
/rpt_family="AluSx"
16532..16646
/rpt_family="MER94"
16855..17145
/rpt_family="L2"
17377..17572
/rpt_family="MIR3"
complement(20036..20330)
/rpt_family="Charliel"
complement(20333..20486)
/rpt_family="Charliel"
20625..20691
/rpt_family="L2"

/clone="RP11-851B10"									
misc_feature	1..68896	/note="assembly_fragment"							
misc_feature	6897..95899	/note="assembly_fragment"							
misc_feature	96000..119268	/note="assembly_fragment"							
misc_feature	119369..141739	/note="assembly_fragment"							
misc_feature	141840..152655	/note="assembly_fragment"							
misc_feature	152756..159466	/note="assembly_fragment"							
misc_feature	159567..163465	/note="assembly_fragment"							
misc_feature	163566..166433	/note="assembly_fragment"							
misc_feature	166534..170112	/note="assembly_fragment"							
misc_feature	170213..172095	/note="assembly_fragment"							
misc_feature	172196..173836	/note="assembly_fragment"							
BASE COUNT 53047 a 34410 c 33807 g 51570 t 1002 others									
ORIGIN									
Query Match 99.4%; Score 503.8; DB 2; Length 173836;									
Best Local Similarity 99.6%; Pred. No. 1.5e-125;									
Matches 505; Conservative 0; Mismatches 2; Indels 0; Gaps 0;									
QY 1	tgagtaatcctaatcagcagattttaaagaacttctcgccaccattataacaata 60								
Db 150274	TGAGTAATCCTAATCAGAGATTTTAAAGAAACTTCTCGCCACCACCAATTAACAATA 150215								
QY 61	tcctctaccaatttgtagtaataatttggctaagtacctaatttttaggtaggcaact 120								
Db 150214	TCCTCTACCAATTTGGTAGTAATAATTTTGTCTAATAGTACTTAATTTTAGGTAGGCACCT 150155								
QY 121	gtgtttatcatatatacctctctctttttgattgtctttctgttttaattggcagcta 180								
Db 150154	GTGTTTATACATATATCCATTCCTCTTTTGTATTGCTCTTCTGTATTATGGGCAGCTA 150095								
QY 181	cctctctggcatctagcagaatgagctgctgcagtttacacaaaagaatgagatcag 240								
Db 150094	CCCTCTGTGCATCTAGCAGAATGAGCTGCTGCAGTTTACACAAAAGAAATGGAGATCAG 150035								
QY 241	agtaacttttggccaccaacgctgctgagaaaattttagtggttactatcatcacacatt 300								
Db 150034	AGTACTTTTGTGCCACCAACGCTGCTGAGAAAATTTGTAGTGTACTATCATCACACATT 149975								
QY 301	acttttatttcacgaatattcaccttcctcggtccctgctgctggccgagaggattgcccta 360								
Db 149974	ACTTTTATTTCATCGAATATTTCCACTTCGCGTCCGTTGGCGGCGAGAGATTGCCGTA 149915								
QY 361	cgcagtctgtacgtatgcagttaactcacagcccttctcgccgaacaaatgttgaggcc 420								
Db 149914	CGCATGCTGTACGTATGTCATGTAACCTACAGGCCCTTCTCGCCCAACATGTTGGAGGC 149855								
QY 421	cttttggagactgtgagacaacagcaacttcagcctgaatactctctttcaattgtgga 480								
Db 149854	CTTTTGGAGCTGTGAGACAACAGTAACCTTCAGCCTGAATCATCTTCTTCAATTGTGGA 149795								
QY 481	caagctgccaagagccttagtaggag 507								
Db 149794	CAAGCTGCCAAGAGCCTTGAGTAGGAG 149768								
RESULT 4									
AC015793/c									
LOCUS	214212 bp	DNA	linear	HTG 13-JUL-2000					
DEFINITION	Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS SEQUENCE SAMPLING.								

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC015793
AC015793.2 GI:7144769
HTG: HTGS_PHASE0.
human.

Homo sapiens.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 214212)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 18, clone RP11-1117D15

Unpublished

2 (bases 1 to 214212)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Balgwin, J., Castle, A., Collangelo, M., Collins, S., Collymore, A.,
Brown, A., Castle, A., Collangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferrelira, P., Fitzhugh, W., Forrest, C., Funk, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 3, 2000 this sequence version replaced gi:6446894.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 695: contig of 695 bp in length
696 795: gap of 100 bp
796 1513: contig of 718 bp in length
1514 1613: gap of 100 bp
1614 2337: contig of 724 bp in length
2338 2437: gap of 100 bp
2438 3161: contig of 724 bp in length
3162 3261: gap of 100 bp
3262 3966: contig of 705 bp in length
3967 4066: gap of 100 bp
4067 4772: contig of 706 bp in length
4773 4872: gap of 100 bp
4873 5564: contig of 692 bp in length
5565 5664: gap of 100 bp
5665 6397: contig of 733 bp in length
6398 6497: gap of 100 bp
6498 7202: contig of 705 bp in length
7203 7302: gap of 100 bp
7303 8028: contig of 726 bp in length
8029 8128: gap of 100 bp

|||||
Db 54033 ATCACACATTTACTTTTATTCATCGAATATTTTCACCTTCCGGTCGTGCGGGCCGAG 53954
Qy 351 gattccctacgcatgtctgtacgtatcatgtaactcacagcccttccctgcccgaaca 410
|||||
Db 53953 GATTGCCGTACGCATGCTGTGAGTATCGATGTAAGTACTCACAGCCCTTTCGTGCCGAACA 53894
Qy 411 tgttgaggcccttttgggaagctgtgcagacacagcaacttcagcgtgaatcatcttt 470
|||||
Db 53993 TGTTCGAGGCGCTTTGGGAAGCTGTGCAGACAAACAGTAACCTTCAGCGTGAATTCATTCCTT 53834
Qy 471 caattgtgacaagctgccaagaggcttgagtaggag 507
|||||
Db 53833 CAATTGTGGCAAGCTGCCAAGAGGCTTGAGTAGGAG 53797
RESULT 5
AX251236/c AX251236 12426 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 204 from Patent WO0168912.
ACCESSION AX251236
VERSION AX251236.1 GI:15984659
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 12426)
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source 1. 12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 71.4%; Score 362.2; DB 6; Length 12426;
Best Local Similarity 82.5%; Pred. No. 2.5e-87;
Matches 415; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Qy 1 tgagtaactcctaactcacaggaattttaaaagaacacctcctgcgccaccattaaacaata 60
Db 8056 TAAATATCTCTAATCACAAAATTTTAAAGAAAATCTTCTAGCCGCCCATTTAAACAATA 7997
Qy 61 tctctaccaaatttgtagtaataattttgctaatagtacctaatttttagtaggcact 120
Db 7996 TCTCCTACCAATTTAATAATAATATTTTACTAATAATACCTAATTTTAAATAAACACT 7937
Qy 121 gtgtttatacatatataccattcctcttttttgatgtgtcttctgtttaaaggcgagcta 180
Db 7936 ATATTTATACATATATCCATCTCTCTCTTTTAAATATATCTTCTATTAATAAACAACTA 7877
Qy 181 cctctctggcatctagcagaatgagctgctgcattttacacaaaaagaatgagatcag 240
Db 7876 CCTCTCTTAACATCTAAACAAAATAAATAACTACTACAATTTACACAAAAAATAAATACTAA 7817
Qy 241 agtacttttttggcaccacaacgctgctgagaaatttgtagtattactatcatcacacatt 300
Db 7816 AATACCTTTTATACCAACCAAGCTATCTAAAAAATTTAATAATTAATTAATCATCACAT 7757
Qy 301 acttttattcatogaattttcaccttccggtccgtgcgtggcgagagattgccgta 360
Db 7756 ACTTTTATTTCATCGAATATTTACCTTCGATCCGATCCGATCAACGAAAGAAAATACCGTA 7697
Qy 361 cgcattgctgtacgtatgcatgtaactcacagcccttccctgcgcgaacatggttgaggagc 420
Db 7696 CGCATATCTATACGTATACATATACTCACACACCCCTTCTTACCCGGAACATATTAANAAC 7637

Qy 421 cttttggagctgtgcagacacagcaacttcagcgtgaatcatcttcttcaattgtgga 480
|||||
Db 7636 CTTTTAAAAACTATACAACAACAATAACTTCAACCTAAATCATTTCTTCAATTATAAA 7577
Qy 481 caagctgccaagaggcttgagta 503
|||||
Db 7576 CAAACTACCAAAAAACCTTAAATA 7554
RESULT 6
AX251235 AX251235 12426 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 203 from Patent WO0168912.
ACCESSION AX251235
VERSION AX251235.1 GI:15984658
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 12426)
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 203 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source 1. 12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3527 a 137 c 2591 g 6171 t
ORIGIN
Query Match 67.8%; Score 343.8; DB 6; Length 12426;
Best Local Similarity 79.9%; Pred. No. 2.3e-82;
Matches 405; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Qy 1 tgagtaactcctaactcacaggaattttaaaagaacacctcctgcgccaccattaaacaata 60
Db 4371 TGAGTAATTTTAAATATAGGATTTTAAAGAAAATTTTTCGCTTATTTTAAATAATA 4430
Qy 61 tctctaccaaatttgtagtaataattttgctaatagtacctaatttttagtaggcact 120
Db 4431 TTTTATTATTAATTTGCTAGTAAATATTTTGTAAATAGTATTTAAATTTTAGTAGGTAT 4490
Qy 121 gtgtttatacatatataccattcctcttttttgatgtgtcttctgtttaaaggcgagcta 180
Db 4491 GTGTTTATATATATATATTTTATTTTATTTTGTATTTTGTATTTTGTATTTTGTATTT 4550
Qy 181 cctctctggcatctagcagaatgagctgctgcattttacacaaaaagaatgagatcag 240
Db 4551 TTTTATTTTGGTATTTAGTAGAATGAGTTGTTGTTAGTTATATAAAAAGAAATGAGATTAG 4610
Qy 241 agtacttttttggcaccacaacgctgctgagaaatttgtagtattactatcatcacacatt 300
Db 4611 AGTATTTTGTGTATTAACGTGTTTGAGAAATTTGAGTGTATTTATTTATTTATATAT 4670
Qy 301 acttttattcatogaattttcaccttccggtccgtgcgtggcgagagattgccgta 360
Db 4671 ATTTTATTTTATCATCGAATATTTTATTTTTCGGTTTTTTCGGTGGCGAGAGATTTCGTA 4730
Qy 361 cgcattgctgtacgtatgcatgtaactcacagcccttccctgcgcgaacatggttgaggagc 420
Db 4731 CTAATGTTTGTACGTATGTAATTTATATATAGTTTATTTTGTTCGAATATGTTGGAGT 4790
Qy 421 cttttggagctgtgcagacacagcaacttcagcgtgaatcatcttcttcaattgtgga 480
Db 4791 TTTTGGGAAGTTGCTAGATAATACTAATTTTGTAGTTTGTGAATTTATTTTAAATTTGGA 4850
Qy 481 caagctgccaagaggcttgagtaggag 507
|||||
Db 4851 TAAAGTTGTTAAGAGGCTTGTAGTAGGAG 4877

AC015793 214212 bp DNA linear HTG 13-JUL-2000
Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
AC015793
HTG: HTGS_PHASE0.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-1117D15
Unpublished
2 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,L., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collimore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J.C., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://fcp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12968 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14589: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp
* 20287 20980: contig of 694 bp in length
* 20981 21080: gap of 100 bp
* 21081 21779: contig of 699 bp in length
* 21780 21879: gap of 100 bp
* 21880 22594: contig of 715 bp in length
* 22595 22694: gap of 100 bp
* 22695 23399: contig of 705 bp in length
* 23400 23499: gap of 100 bp
* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length
* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp


```
misc_feature complement(23173..23244)
misc_feature /note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"
23184..23261
repeat_region /note="similar to EST AA366493 (NID:g2018832)"
complement(27159..27188)
repeat_region /rpt_family="L1"
28018..28045
repeat_region /rpt_family="L1"
complement(28082..28155)
repeat_region /rpt_family="L1"
28086..28175
repeat_region /rpt_family="L1"
28200..28284
repeat_region /rpt_family="ALU"
complement(29362..29403)
repeat_region /rpt_family="L1"
29513..29594
repeat_region /rpt_family="L1"
29943..30142
repeat_region /rpt_family="L1"
30183..30377
repeat_region /rpt_family="L1"
30980..31186
repeat_region /rpt_family="L1"
31384..31422
repeat_region /rpt_family="L1"
31441..31480
repeat_region /rpt_family="L1"
33572..33606
repeat_region /rpt_family="L1"
34091..34382
repeat_region /rpt_family="ALU"
complement(34850..35137)
repeat_region /rpt_family="ALU"
35955..36007
misc_feature /note="similar to EST AA425675 (NID:g2106395) zw47b05.s1"
complement(38607..39020)
repeat_region /rpt_family="L1"
39165..39186
repeat_region /rpt_family="L1"
complement(41821..41959)
repeat_region /rpt_family="L1"
41957..43047
repeat_region /rpt_family="L1"
43072..43358
repeat_region /rpt_family="L1"
43364..43656
repeat_region /rpt_family="ALU"
43682..44482
repeat_region /rpt_family="L1"
complement(44490..44609)
repeat_region /rpt_family="L1"
44565..46946
repeat_region /rpt_family="L1"
complement(48709..48999)
repeat_region /rpt_family="ALU"
49881..49901
repeat_region /rpt_family="L1"
51668..56075
repeat_region /rpt_family="L1"
complement(52387..52447)
misc_feature /note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"
52387..52466
misc_feature /note="similar to EST AA366493 (NID:g2018832)"
complement(52387..52480)
repeat_region /note="similar to EST AA48997 (NID:g2218599) aa54c11.s1"
56076..56419
repeat_region /rpt_family="MER"
56420..56942
repeat_region /rpt_family="L1"
complement(57178..57226)
repeat_region /rpt_family="ALU"
59193..59213
```

```
repeat_region /rpt_family="L1"
complement(60100..61153)
repeat_region /rpt_family="MER"
63986..64015
repeat_region /rpt_family="L1"
64113..64253
repeat_region /rpt_family="ALU"
complement(65648..65939)
repeat_region /rpt_family="ALU"
66049..66338

Query Match 8.8%; Score 44.4; DB 9; Length 170136;
Best Local Similarity 46.7%; Pred. No. 0.25;
Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 1 tgaagtaacctcaatcacaggatttttaaaagaacttctgcgcacaccattaaacaata 60
Db 62665 TAAGTAATCTCACTGACATATTTCTTAGAAAAAAGATACAAATATTTTCCCTTGC 62724
Qy 61 tctctaccatattggtagtaataattttgctaaatagtagctaaatttttagtaggcact 120
Db 62725 ACTCTATAGTGTCTCTCAAAATATTTTAAAGTCATCTTAGTTTATTTTCTTATA 62784
Qy 121 gtgtttatcatatccattcctctttttttgttgattgtttcttcttaattgggcagcta 180
Db 62785 TTCACAATGCTTAATAAGATTTTGTAGACTCTCATGCTGAATGATGTAATGTGTAGATT 62844
Qy 181 cctctcttggeatctagcagaatgagctgctgcaatttaccacaaaagaatgagatcag 240
Db 62845 CAACGTGCTAGTGTAAACACACAGCTATGCTCATCTCTTTGTGAAGCTTAAATATATG 62904
Qy 241 agtactttttgtgcacacacgtgtctgagaattgttagtatttagtattcatcacacatt 300
Db 62905 TCTGTCTTTTATCCAATTGAGTCCCTCGTAACCTACCTGCTGTGTAACAATATCACAGTTT 62964
Qy 301 ac 302
Db 62965 CC 62966

RESULT 10
AC084396/c
LOCUS AC084396 225053 bp DNA linear HTG 05-APR-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-134N3, WORKING DRAFT
SEQUENCE, 61 unordered pieces.
ACCESSION AC084396
VERSION AC084396.3 GI:13549348
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 225053)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Apr 5, 2001 this sequence version replaced gi:12863223.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0134N03
----- Summary Statistics -----
Sequencing vector: M13; 91%
Sequencing vector: plasmid; 4%
```

Chemistry: Dye-terminator ET; 91% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184917 bases at least Q40
Consensus quality: 196549 bases at least Q30
Consensus quality: 203318 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 219053; sum-of-contigs
Quality coverage: 3.56 in Q20 bases; agarose-fp
Quality coverage: 2.91 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 134126	135771:	contig of 1646 bp in length
* 135772	135871:	gap of unknown length
* 135872	137168:	contig of 1297 bp in length
* 137169	137268:	gap of unknown length
* 137269	138776:	contig of 1508 bp in length
* 138777	138876:	gap of unknown length
* 138877	140630:	contig of 1754 bp in length
* 140631	140730:	gap of unknown length
* 140731	143390:	contig of 2660 bp in length
* 143391	143490:	gap of unknown length
* 143491	144885:	contig of 1395 bp in length
* 144886	144985:	gap of unknown length
* 144986	146277:	contig of 1292 bp in length
* 146278	146377:	gap of unknown length
* 146378	148790:	contig of 2413 bp in length
* 148791	148890:	gap of unknown length
* 148891	151618:	contig of 2728 bp in length
* 151619	151718:	gap of unknown length
* 151719	153528:	contig of 1810 bp in length
* 153529	153628:	gap of unknown length
* 153629	155632:	contig of 2004 bp in length
* 155633	155732:	gap of unknown length
* 155733	158353:	contig of 2621 bp in length
* 158354	158453:	gap of unknown length
* 158454	161521:	contig of 3068 bp in length
* 161522	161621:	gap of unknown length
* 161622	163929:	contig of 2308 bp in length
* 163930	164029:	gap of unknown length
* 164030	165640:	contig of 1611 bp in length
* 165641	165740:	gap of unknown length
* 165741	168699:	contig of 2959 bp in length
* 168699	168799:	gap of unknown length
* 168799	169971:	contig of 1172 bp in length
* 169972	170071:	gap of unknown length
* 170072	171383:	contig of 1312 bp in length
* 171384	171483:	gap of unknown length
* 171484	174613:	contig of 3130 bp in length
* 174614	174713:	gap of unknown length
* 174714	178023:	contig of 3310 bp in length
* 178024	178123:	gap of unknown length
* 178124	179809:	contig of 1686 bp in length
* 179810	179909:	gap of unknown length
* 179910	182166:	contig of 2257 bp in length
* 182167	182266:	gap of unknown length
* 182267	184778:	contig of 2512 bp in length
* 184779	184878:	gap of unknown length
* 184879	187289:	contig of 2411 bp in length
* 187290	187389:	gap of unknown length
* 187390	190025:	contig of 2636 bp in length
* 190026	190125:	gap of unknown length
* 190126	193628:	contig of 3503 bp in length
* 193629	193728:	gap of unknown length
* 193729	196385:	contig of 2657 bp in length
* 196386	196485:	gap of unknown length
* 196486	199287:	contig of 2802 bp in length
* 199288	199387:	gap of unknown length
* 199388	202073:	contig of 2686 bp in length
* 202074	202173:	gap of unknown length
* 202174	205921:	contig of 3748 bp in length
* 205922	206021:	gap of unknown length
* 206022	211281:	contig of 5260 bp in length
* 211282	211381:	gap of unknown length
* 211382	215785:	contig of 4404 bp in length
* 215786	215885:	gap of unknown length
* 215886	220968:	contig of 5083 bp in length
* 220969	221088:	gap of unknown length
* 221089	225053:	contig of 3985 bp in length
* 221069	Location/Qualifiers	
	1. .225053	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="UNK"	
	/clone="RP11-134N3"	

FEATURES

source

134026

Campopiano,A., Castle,A., Choepei,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,K., Landers,T., Lehocsky,J.,
Levine,R., Liou,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,C., McPheeters,R.,
McDrirm,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Ollivari,T.M., Oliver,J., Peterson,K., Plierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zalcun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 6, 2000 this sequence version replaced gi:7239678.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----
Center project name: L7901
Center clone name: 243.M.7
----- Summary Statistics -----
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156282 bases at least Q40
Consensus quality: 160561 bases at least Q30
Consensus quality: 162162 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 163139; sum-of-ctngs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctngs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3738: contig of 3738 bp in length
* 3739 3838: gap of 100 bp
* 3839 6584: contig of 2746 bp in length
* 6585 6684: gap of 100 bp
* 6685 13062: contig of 6378 bp in length
* 13063 13162: gap of 100 bp
* 13163 21786: contig of 8624 bp in length
* 21787 21886: gap of 100 bp
* 21887 30518: contig of 8632 bp in length
* 30519 30618: gap of 100 bp
* 30619 42483: contig of 11865 bp in length
* 42484 42583: gap of 100 bp
* 42584 54609: contig of 12026 bp in length
* 54610 54709: gap of 100 bp
* 54710 68804: contig of 14095 bp in length
* 68805 68904: gap of 100 bp
* 68905 84136: contig of 15232 bp in length
* 84137 84236: gap of 100 bp
* 84237 99346: contig of 15110 bp in length
* 99347 99446: gap of 100 bp

FEATURES	source	99447 116848: contig of 17402 bp in length	
		* 116849 116948: gap of 100 bp	
		* 116949 139439: contig of 22491 bp in length	
		* 139440 139539: gap of 100 bp	
		* 139540 164339: contig of 24800 bp in length.	
	Location/Qualifiers		
	1. .164339	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="11"	
		/map="11"	
misc_feature		/clone_lib="RPC1-11 Human Male BAC"	
	1. .3738	/note="assembly_fragment"	
	3839. .6584	/note="assembly_fragment"	
	6685. .13062	/note="assembly_fragment"	
	13163. .21786	/note="assembly_fragment"	
	21887. .30518	/note="assembly_fragment"	
	30619. .42483	/note="assembly_fragment"	
	42584. .54609	/note="assembly_fragment"	
	clone_end:SP6	vector_side:left	
	54710. .68804	/note="assembly_fragment"	
misc_feature	68905. .84136	/note="assembly_fragment"	
	84237. .99346	/note="assembly_fragment"	
	clone_end:T7	vector_side:left	
	99447. .116848	/note="assembly_fragment"	
	118949. .139439	/note="assembly_fragment"	
	139540. .164339	/note="assembly_fragment"	
	52600 a 30109 c 28400 g 52030 t 1200 others		
	BASE COUNT		
	ORIGIN		
Query Match	Best Local Similarity	8.4%; Score 42.6; DB 2; Length 164339;	
	Mismatches	99; Conservative 0; Mismatches 94; Indels 0; Gaps 0;	
	QY 120	tgtgtttatcatatccattcctctcttttggattgtcttctgttttaattggtgagct 179	
	DB 100379	TGTATATATATATATATATATATATGATTTTAAATGTGAAACGTTTCTTAGACGTA 100320	
	QY 180	acctctctggcatcagagaatgagctgctgcagtttacacaaaagaatgagatca 239	
	DB 100319	CACACTCTATGTGTAACCAAAATTAATTATTCATTTTAAAGATCCAGAACTGTAATA 100260	
	QY 240	gagtcatttttggccacaacgctgtctgagaaattgttagttactatcatcacat 299	
	DB 100259	GCTACCATTTTGTGGACACTTGCTATTTGCTATATTTATTTGCTATATGTTGACAGATAT 100200	
	QY 300	tactttttattca 312	
	DB 100199	AAATATTATTACA 100187	
RESULT 12			
AC084783		166982 bp DNA	linear PRI 13-JAN-2001
LOCUS			
DEFINITION	Homo sapiens chromosome 15 clone RP11-5N19 map 15q21.3, complete sequence.		

AC084783 AC015659

AC084783.2 GI:12203289

HTG.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 166982)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.

Sequencing of human chromosome 15 D15S146-D15S117 region

Unpublished

2 (bases 1 to 166982)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.

Direct Submission

Submitted (16-NOV-2000) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

3 (bases 1 to 166982)

Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A., Nesbitt, R., Traicoff, R. and Hood, L.

Direct Submission

Submitted (13-JAN-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA

On Jan 13, 2001 this sequence version replaced gi:11181781.

----- Genome Center

Center: Multimegabase Sequencing Center

Center code: UWMSC

Web site: http://chroma.mbt.washington.edu/msg_www

Contact: leerowen@systemsbiology.org

Drafting center: WIBR

----- Summary Statistics

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 90% of reads

Chemistry: Dye-primer Big Dye; 10% of reads

Assembly program: Phrap; version 0.990399

This clone was finished using overlapping sequence from AC036163 [drafting center=WIBR] and AC068726 [drafting center = UWMSC]

----- Location/Qualifiers

1. .166982

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="15"

/map="15q21.3"

/clone="RPC1 human BAC library 11"

/clone="RP11-5N19"

/note="This clone overlaps RP11-241D5 AC036163 and RP11-844G18 AC068726. Data from these BACs were combined and the consensus sequence was derived from 5N19 to the extent possible."

misc_feature

1. .34304

/note="overlap with RP11-241D5, AC036163"

47770. .47790

/note="low quality data"

77830. .78230

/note="low quality data"

82770. .82790

/note="low quality data"

104080. .104130

/note="low quality data"

159310. .166982

/note="overlap with RP11-844G16, AC068726"

BASE COUNT

52301 a 32365 c 31886 g 50430 t

ORIGIN

```
Query Match      8.4%; Score 42.4; DB 9; Length 166982;
Best Local Similarity 54.5%; Pred. No. 0.87;
Matches 85; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 72 ttgtgtagtaaatatttgctaatagtagctacctaatttttaggtaggcactgtgtttatata 131
    || ||| ||||| || ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135566 TTGTTAGGGGATATATGTCGATATTTACTTTACTTTTGTGATAATACACTGTTGTAATA 135625
    || ||| ||||| || ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 132 tatatactctctctcttttggatgctctcttctgattggaatggcgagctacctctcttggc 191
    ||||| || ||| ||| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135626 ACGATCCAGTAATAATTTACTGAGTGCTCTTCATGAAGAAGACAGTGTAAAGTACTAGGG 135685
    || ||| || ||| ||| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 192 atctagcagaatgagctgctgcagtttacacaaaa 227
    || ||| || ||| ||| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135686 ATATAGAGACATAAGTCCCGAGCCCTTACTTTCACAAA 135721
    || ||| || ||| ||| ||||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AB022222/c
LOCUS      AB022222      22601 bp      DNA      linear      PLN 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MUD12.
ACCESSION  AB022222 BA000015
VERSION    AB022222.1 GI:4159711
KEYWORDS
SOURCE     Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
           clone:MUD12.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
            1 (sites)
            Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H.
            and Tabata,S.
            Structural analysis of Arabidopsis thaliana chromosome 5. X.
            Sequence features of the regions of 3,076,755 bp covered by sixty
            P1 and TAC clones
            DNA Res. 7 (1), 31-63 (2000)
            20181125
            2 (bases 1 to 22601)
            Nakamura,Y.
            Direct Submission
            Submitted (08-JAN-1999) Yasukazu Nakamura, Kazusa DNA Research
            Institute, Department of Plant Gene Research; 1532-3, Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
            Tel:81-438-52-3935, Fax:81-438-52-3934)
            Address for correspondence: kaos@kazusa.or.jp
            For the latest information on annotation of this clone, please see
            http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MUD12
            Genes with similarity to proteins in the databases are described in
            'product' or 'note' qualifiers. Genes that have no significant
            protein similarity are described as 'unknown protein'.
            The software programs used to predict genes include: Grail
            (Informatics Group, Oak Ridge National Laboratory,
            http://compbio.ornl.gov/grail-1.3/),
            GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
            NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
            Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
            SplicePredictor (Volker Brendel, Stanford University,
            http://grm1n1n1.zoool.iastate.edu/cgi-bin/sp.cgi).
            Genes encoding tRNAs are predicted by tRNAscan-SE
            (Sean Eddy, Washington University School of Medicine, St. Louis,
            http://genome.wustl.edu/eddy/tRNAscan-SE/).
            This sequence may not be the entire insert of this clone. It may be
            shorter because we remove overlaps between neighboring submissions.
            The 5' clone is MWH19 and the 3' clone is MSN9.
            Location/Qualifiers
                1. :22601
                   /organism="Arabidopsis thaliana"
                   /strain="Columbia"
                   /db_xref="taxon:3702"
                   /chromosome="5"
                   /clone="MUD12"
                   /clone_lib="Mitsui P1"

FEATURES             source
CDS
```

```
join(1431..2080,2638..3487)
/note="gene_id:MUD12.1"
/codon_start=1
/evidence=not_experimental
/product="protein kinase"
/protein_id="BAA97351.1"
/db_xref="GI:8843803"
/translation="MSKQTPNLESQDCSSSNYNANASPVSGSASFKTSNGSEDEV
LNNSISLFCSSNSVSSEANLEKTOFSDANEANFKRVFAPSKPHKGNLDRDAIQNVK
CKNEDLGLGHFRLLKILGCGDIGSVYLAELREMGCFEAMKMDKMLKVLVRAQ
TEREILGLDHPFLPTLYSHPETEKFSCLLMEFCGGDLHLRQKQKGFSELAARF
YASVLLALALEYLLHMGVYVYRDKPENVMVRDEGHIMLSDEFDLSLQSPVSPFLI
PSCHIASYQPPCIDPCSPKACIQPCFKPFLNKKPKAKTERKAGSDSLPMLTA
EPTAARSMFSVFGTHYLAETIRGDGHGSSVDMWTFGFLYLLTGTGTFPKGNRRE
LFNVGQPLKFPFEGSISFAAKDLIRGLLTKDKPKRLGFKKGATGTEIKOHPFN
IRSTTEPEIKPIDLSILNETLKSSVQOQGHKSKQSDSSSGPYLDFEF"
complement(5117..5461)
/note="gene_id:MUD12.2
similar to unknown protein
sp|O24415"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97352.1"
/db_xref="GI:8843804"
/translation="MKVVAAYLLAKLSGNENPSVADLKKIVESVGAETDQEKIDLF
LKRDVTELIAGVREKMAALSGGGAVAVASGGGGAAPAAEPASVESKKEEKEE
SEDDGGMMSLFD"
complement(join(7287..7481,7777..8253,8519..8755))
/note="gb|AAD23631.1
gene_id:MUD12.3
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97353.1"
/db_xref="GI:8843805"
/translation="METRKLVSLSGSSIDSTISPLDELLSHLSLSELPKRAASTSIL
KRWTLFPLMNLHCASLDDTDLDPERQTEEEYVSVFVGLLGLTIFPCLRVSSST
LSNHAICDVMGCPQLEELYIHIIYEDRDEEFNIHTAPHYIAHNTKKLTCVYNNDY
QAPRLISYTPNLVLDYSDYLTCLYNSENHFNDLLEARLDAFARAGMGDEHTCL
KIMNSITNVQILHLSCTVEVEIQEYKGRLELNQVKCFLREMLEEAVKNTSDEI
ENKQLTNDLALLPKRSSCKNIHL"
join(9792..10258,10626..11451,11535..11843,11943..12350,
12638..13816,14494..14763)
/note="gene_id:MUD12.4"
/codon_start=1
/evidence=not_experimental
/product="disease resistance protein-like"
/protein_id="BAA97354.1"
/db_xref="GI:8843806"
/translation="MAYSSSCNNVYDVFPFSGEDVRKTKFLSHQLVLDRKLITSPK
DNTERSQSIAPELVQGIKDSRIAIVIFSKNYASSMCLNELLETVSCDKGOLVIP
VFYALDPTHVRKQTGDGMAFERCLANKTEDEKNLRVALTHVANILGYHSAQCRANP
DDYNKJLHQLTETLSTIGKONIKIDHILGALGERLKHQKVLFFIDDDQVYVNLAG
QIQWFGSGSRIIVYVNDKLLSHIGIENIQVCLPSKELALEMLCRVAFRQNTPPDG
KLAVEVYVHAGILPLGLNVLSYLRGNKRWMDMLPLRKLDGDKGTQAKRLGYDG
LDNKKDAIFRHIACLNFENKNDIRLLADSLNFIENLVKSLVNVRSNIVEM
HCLQEMGREIVRAQSNAGREFLMDTDCVLDNDNIGTKMGLISLVDIEDHLEL
NVHEKATQGMENLFLNITKALMSGOKIRLHPENEDYLPKLLKLCWDKYPWRCLP
SSFRPENVLKLMOESELKELWEGVSLTCLKMDMLEKSKNKEIPDLSMATNLKTLN
LKYCSSLVKISSIONLKLTKLNMEGCTNLETIPAGINUKSLRHLDPDUGCSRLRMFP
DINSISVILFDDSTIEEFLNHLKLFDLSMQOMNSEKLMWEGVQPTCLUMKMLSP
LKNKNTLYLSDIPSLVPCGIQNLKLMELSIRRCNLTSLPTGANFKLYDLVLS
GCKSLRSPDISSTISCLNRTGIEEVPVSIENFVRLTYLTMLECNKLVVSLNIFK
LKLHLDADFDCGTLTEVSMCKNTISVAATADNIOPKLLVSEASSSLCVOXSVREI
NCFKLQDEALQQEPVKSLILGEEVPAENHRATGNSLVIPLVPTSISLDFLGFR
CALVDVVRKAMPGRVDIQVSCFRGSLKHFDSADSHSLVAFHKASHLLIFDCRFAL
NDSNPLNYAHMDITFHLTDSVKINACGIRFSEGVSPQKLNKPSIIVHVCDES
KLDNDRFAHETHEHTEGCGSDNVETVYSRKRNKRNKKNKKNKLNKRNKRTSLKN
KGLNPKRKTSLNVLYPLADHYSPQISQLISLEPHVSKVRERFVLLSRNRSQ
NRIIR"
complement(join(15394..15462,15748..15812,16583..16733))
/note="gene_id:MUD12.5
ref|NP_009841.1
```


Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczyk, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.	24451	25766:	contig of 1316 bp in length
Direct Submission	25767	27235:	gap of unknown length
Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	27236	29226:	gap of unknown length
All repeats were identified using RepeatMasker:	29227	30961:	gap of unknown length
Smit, A.F.A. & Green, P. (1996-1997)	30962	32608:	gap of unknown length
http://ftp.genome.washington.edu/RM/RepeatMasker.html	32609	34331:	gap of unknown length
----- Genome Center	34332	36533:	gap of unknown length
Center: Whitehead Institute/ MIT Center for Genome Research	36534	39299:	gap of unknown length
Center code: WIBR	39300	41451:	gap of unknown length
Web site: http://www-seq.wi.mit.edu	41452	43295:	gap of unknown length
Contact: sequence.submissions@genome.wi.mit.edu	43296	45678:	gap of unknown length
----- Project Information	45679	48551:	gap of unknown length
Center project name: L5217	48552	51186:	gap of unknown length
Center clone name: 335_E_2	51187	53207:	gap of unknown length
-----	53208	55697:	gap of unknown length
* NOTE: This is a 'working draft' sequence. It currently	55698	58328:	gap of unknown length
* consists of 58 contigs. The true order of the pieces	58329	59424:	gap of unknown length
* is not known and their order in this sequence record is	59425	62209:	gap of unknown length
* arbitrary. Gaps between the contigs are represented as	62210	64795:	gap of unknown length
* runs of N, but the exact sizes of the gaps are unknown.	64796	67972:	gap of unknown length
* This record will be updated with the finished sequence	67973	71707:	gap of unknown length
* as soon as it is available and the accession number will	71708	74928:	gap of unknown length
* be preserved.	74929	78667:	gap of unknown length
1	78668	82959:	gap of unknown length
1169: contig of 1169 bp in length	82960	88848:	gap of unknown length
gap of unknown length	88849	93807:	gap of unknown length
1170	93808	98630:	gap of unknown length
2744: contig of 1575 bp in length	98631	106830:	gap of unknown length
gap of unknown length	106831	114298:	gap of unknown length
2745	114299	121722:	gap of unknown length
3891: contig of 1147 bp in length	121723	128019:	gap of unknown length
gap of unknown length	128020	136332:	gap of unknown length
3892	136333	14730:	gap of unknown length
4958: contig of 1067 bp in length	14731	153111:	gap of unknown length
gap of unknown length	153112	162589:	gap of unknown length
4959	162590	171983:	gap of unknown length
6042: contig of 1084 bp in length	171984	181733:	gap of unknown length
gap of unknown length			
6043			
7112: contig of 1070 bp in length			
gap of unknown length			
7113			
8242: contig of 1130 bp in length			
gap of unknown length			
8243			
9441: contig of 1198 bp in length			
gap of unknown length			
9441			
10503: contig of 1063 bp in length			
gap of unknown length			
10504			
11766: contig of 1263 bp in length			
gap of unknown length			
11767			
12798: contig of 1032 bp in length			
gap of unknown length			
12799			
14299: contig of 1501 bp in length			
gap of unknown length			
14300			
15314: contig of 1015 bp in length			
gap of unknown length			
15315			
16978: contig of 1664 bp in length			
gap of unknown length			
16979			
18049: contig of 1071 bp in length			
gap of unknown length			
18050			
19597: contig of 1548 bp in length			
gap of unknown length			
19598			
21003: contig of 1406 bp in length			
gap of unknown length			
21004			
22126: contig of 1123 bp in length			
gap of unknown length			
22127			
23502: contig of 1376 bp in length			
gap of unknown length			
23503			
24450: contig of 948 bp in length			
gap of unknown length			

TITLE
JOURNAL
COMMENT

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:30 : Search time 403.26 Seconds
(without alignments)
2158.596 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507
Sequence: 1 tgaatactactaaccacagg.....ccaagagcgttagtagagag 507

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507	100.0	1141	18	AAT89000 Human maspin Promo
2	362.2	71.4	12426	22	AAS46482 Tumour suppressor
3	343.8	67.8	12426	22	AAS46481 Tumour suppressor
4	38.2	7.5	1239	22	AAS68963 Human immune/haema
5	38.2	7.5	13537	22	AAS68964 Human immune/haema
6	38	7.5	56153	22	AAS46793 Tumour suppressor
7	37.4	7.4	6063	22	AAS46337 Tumour suppressor
8	37	7.3	7143	21	AAV0250 Plasmodium falcipar
9	36.8	7.3	513445	22	AAI61373 Soybean 318013 reg

10	36.6	7.2	8964	24	ABL33534 Human immune syste
11	36.6	7.2	8964	24	AAS61276 Human gene regulat
12	36.4	7.2	4421	22	AAS85974 Human immune/haema
13	36.4	7.2	61020	22	AAS46787 Tumour suppressor
14	36.2	7.1	2234	17	AAT09022 Arabidopsis thalia
15	36.2	7.1	6309	24	ABL32304 Human immune syste
16	35.8	7.1	6309	24	ABL33615 Human immune syste
17	35.8	7.1	8123	24	ABL34227 Human immune syste
18	35.6	7.0	952	19	AAK14118 H. pylori GHPO 122
19	35.6	7.0	9293	22	AAS45320 Chemically pretrea
20	35.4	7.0	7403	22	AAS46804 Tumour suppressor
21	35.4	7.0	7403	24	ABL34217 Human immune syste
22	35.4	7.0	56153	22	AAS46794 Tumour suppressor
23	35.2	6.9	11047	22	AAS45480 Chemically pretrea
24	35.2	6.9	11047	24	ABL33985 Human immune syste
25	35.2	6.9	28091	22	ABAI5375 Human nervous syste
26	35.2	6.9	28091	22	AAK77432 Human immune/haema
27	35.2	6.9	28120	22	ABAI5379 Human nervous syst
28	35.2	6.9	28120	22	AAK77434 Human immune/haema
29	35.2	6.9	114793	22	AAO08215 Human genome from
30	35	6.9	15592	22	AAS46454 Tumour suppressor
31	35	6.9	15592	24	ABL33327 Human immune syste
32	35	6.9	16509	24	ABL33321 Human immune syste
33	34.8	6.9	1494	20	AAK20170 Enterococcus faeca
34	34.8	6.9	6134	24	ABL33183 Human immune syste
35	34.8	6.9	6259	20	AAK13066 Enterococcus faeca
36	34.8	6.9	7069	22	AAS46654 Tumour suppressor
37	34.8	6.9	7069	24	ABL33353 Human immune syste
38	34.8	6.9	7069	24	AAS61220 Human gene regulat
39	34.6	6.8	370	22	AAS4576 Human polynucleotl
40	34.6	6.8	890	22	ABL15375 Human breast cance
41	34.6	6.8	1192	22	ABAI9853 Human nervous syst
42	34.6	6.8	6136	24	AAS61297 Human gene regulat
43	34.6	6.8	18028	23	ABL33284 Drosophila melanog
44	34.6	6.8	25955	22	ABAI9852 Human nervous syst
45	34.4	6.8	834	22	AAH03576 Human cDNA clone (

ALIGNMENTS

RESULT 1	
AAT89000	standard; cDNA; 1141 BP.
AC	AA789000:
XX	22-APR-1998 (first entry)
XX	Human maspin promoter and partial coding cDNA.
DE	Human maspin promoter and partial coding cDNA.
XX	Masspin: serpin; mammary epithelial cell; human; promoter; malignant;
KW	tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.
OS	Homo sapiens.
XX	
FT	key
FT	promoter
FT	Location/Qualifiers
FT	1..1141
FT	/*tag- a
FT	/note- "maspin promoter region"
FT	443..449
FT	/*tag- b
FT	/note- "Putative regulatory element AP2"
FT	451..457
FT	/*tag- c
FT	/note- "Putative regulatory element AP1"
FT	485..490
FT	/*tag- d
FT	/note- "Putative regulatory element Ets"
FT	684..691
FT	/*tag- e
FT	/note- "Putative regulatory element HRE"
FT	846..851
FT	misc_signal

```

FT      /tag= f
FT      /note= "Putative regulatory element Ets"
FT      misc_signal
FT      907..913
FT      /tag= g
FT      /note= "Putative regulatory element AP1"
FT      prim_transcript 958..1141
FT      /tag= h
FT      /note= "partial"
XX
XX      WO9736179-A1.
XX
XX      02-OCT-1997.
XX
XX      28-MAR-1997; 97WO-US05186.
XX
XX      28-MAR-1996; 96US-0014368.
XX
XX      (DAND ) DANA FARBER CANCER INST INC.
XX      (DAND ) DANA FARBER CANCER INST.
XX      (PARD/) PARDEE A.
XX
XX      Sagar R, Zhang M;
XX      WPI; 1997-489785/45.
XX
XX      Maspin gene promoter fragment - used to identify compounds for
XX      treatment of prostate or breast cancer
XX
XX      Claim 4; Fig 3; 51pp; English.
XX
XX      This sequence encodes the human maspin promoter region, including the
XX      transcription start site. Maspin is a serpin which is expressed in
XX      mammary epithelial cells. Its expression in these cells decreases
XX      with increasing malignancy and is lost in during metastasis. Maspin
XX      protein is also known to inhibit the mobility of tumour cells. This gene
XX      can be used in method for screening compounds to identify candidate
XX      compounds for the treatment of prostate cancer, or breast cancer. It can
XX      also be used to identify compounds that increase the expression of
XX      maspin, and for detecting the presence of metastatic prostate epithelial
XX      cells.
XX
XX      Sequence 1141 BP; 261 A; 292 C; 277 G; 311 T; 0 other;
XX

```

```

Query Match      100.0%; Score 507; DB 18; Length 1141;
Best Local Similarity 100.0%; Pred. NO.2.9e-135;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 tgaatacctcaatcacagagatttaaaagaactctctgcccacccattaaacata 60
DB      451 tgaatacctcaatcacagagatttaaaagaactctctgcccacccattaaacata 510
OY      61 tctctaccgaatttggtgtaataatttggtaataagactaattttagtgagact 120
DB      511 tctctaccgaatttggtgtaataatttggtaataagactaattttagtgagact 570
OY      121 ggtttatacatatatacatcctctcttttgatgtgcttcgtgttaatgagcgacta 180
DB      571 ggtttatacatatatacatcctctcttttgatgtgcttcgtgttaatgagcgacta 630
OY      181 cctctcttgacatctagcagaatgagctgctgacgtttacacaaagaatgagagatag 240
DB      631 cctctcttgacatctagcagaatgagctgctgacgtttacacaaagaatgagagatag 690
OY      241 agtaacttttggcacaacagctgctgagaaattttagtggtactatcacacatt 300
DB      691 agtaacttttggcacaacagctgctgagaaattttagtggtactatcacacatt 750
OY      301 acttttaattcatgaataattcaccttcgctgctgctgctgctgctgctgctgctgct 360
DB      751 acttttaattcatgaataattcaccttcgctgctgctgctgctgctgctgctgctgct 810
OY      361 cgcattgtctgtaagatgcatgtaactcacagccctctctgcccgaacatgttgagac 420

```

```

DB      811 cgcattgtctgtaagatgcatgtaactcacagccctctctgcccgaacatgttgagac 870
OY      421 ctttgaagctgtgcaacaacagcaacttcacaggaatcatctcttcaatgtgga 480
DB      871 ctttgaagctgtgcaacaacagcaacttcacaggaatcatctcttcaatgtgga 930
OY      481 caagctgcccaagagagctgtagtagag 507
DB      931 caagctgcccaagagagctgtagtagag 957

RESULT 2
AAS46482/C
ID      AAS46482 standard; DNA; 12426 BP.
XX
XX      AAS46482;
AC
XX      18-DEC-2001 (first entry)
DT
XX      Tumour suppressor gene derived chemically modified sequence #204.
DE
XX      Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX      cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX      cytosine methylation; ds.
XX      Homo sapiens.
XX      WO200168912-A2.
XX
XX      20-SEP-2001.
PD
XX      15-MAR-2001; 2001WO-EP02955.
XX
XX      15-MAR-2000; 2000DE-1013847.
XX      06-APR-2000; 2000DE-1019058.
XX      07-APR-2000; 2000DE-1019173.
XX      30-JUN-2000; 2000DE-1032529.
XX      01-SEP-2000; 2000DE-1043826.
XX
XX      (EPIC-) EPIGENOMICS AG.
XX
XX      Olek A, Piepenbrock C, Berlin K;
XX
XX      WPI; 2001-602752/68.
XX
XX      Fragments of chemically modified genes associated with tumour suppressor
XX      genes and oncogenes, useful in designing primers and probes for
XX      analysing diseases associated with cytosine methylation state e.g.
XX      cancer
XX
XX      Claim 1; SEQ ID NO 204; 27pp; English.
XX
XX      The invention relates to a nucleic acid comprising a sequence of 18
XX      bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX      bisulphite, of genes associated with tumour suppression and
XX      oncogenes having a sequence taken from 536 (actually 533 since
XX      numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX      (5S) and sequences complementary to (5S). The nucleic acid may be a
XX      peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX      form part of a set of probes for detecting the cytosine methylation state
XX      and/or single nucleotide polymorphisms and also to be used in an
XX      array for analysing diseases associated with CpG dinucleotides e.g.
XX      cancers and tumours. The probes can also be used in a method for
XX      ascertaining genetic and/or epigenetic parameters for the diagnosis
XX      and/or therapy of existing diseases or the predisposition to specific
XX      diseases, by analysing cytosine methylations. The parameters may be
XX      compared to another set of genetic and/or epigenetic parameters, the
XX      differences serving as basis for diagnosis and/or prognosis events which
XX      are disadvantageous to patients. The present sequence is one of the
XX      533 genomic sequences derived from tumour suppressor genes and
XX      oncogenes. Sequences with even numbered Seq ID numbers are the
XX      complementary sequence of the corresponding odd numbered sequence (e.g.

```


CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

Query Match 71.4%; Score 362.2; DB 22; Length 12426;
 Best Local Similarity 82.5%; Pred. No. 2.1e-93;
 Matches 415; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

OY 1 tgaagtaatccatcacagagatttaaaagaactctctgcgcacccatcaacaata 60
 DB 8056 TAAATATCTATACACAAATTTTAAAAAAACTCTCTACGCCACCATTAACATA 7997
 OY 61 tctccaccaatttgtagtaataatttgctaagtaacctaatlttagtaggagcaat 120
 DB 7996 TCTCTACCAATTTATATATAATATTACTAATATACCTAATTTTAAATAAACACT 7937
 OY 121 ggtttaacataatcaatccctctctttttgattgcttcgtttaatgagcaacta 180
 DB 7936 AATATTATACATATATCCATTCCTTTTAAATATCTTTTCTAATTAACAACATA 7877
 OY 181 cctctctggacatcagagatgagctgctgcagtttaacaaaaagaatgagatcag 240
 DB 7876 CTTCTTTTACATCTAACAAATTAACATCTACAAATTTACACAAAAAATTAATCA 7817
 OY 241 agtaattttgtgcacacagctgctcgaagaatttgtagtctactatcacacacatt 300
 DB 7816 AATACTTTTATACACCAAGCTATCTAAAAATTTATATATCTATCATCACACAT 7757
 OY 301 actttattcatcgaataattccaccctcgctcgtcgtagcgagagagattgcgta 360
 DB 7756 ACTTTTATTCATTCGAATATTTCCACCTTCGATCCCTACTAAACGAAAAATTCACGTA 7697
 OY 361 gcgactctgtagcgtatgcatgtaactcacagcccttcctcgcgaagaattgtaggc 420
 DB 7696 CGCATATCTATACGTRATACATATTAACCTCAACCCCTTCTTCCGACATTTTAAAC 7637
 OY 421 ctttggaaagctgtagcagaacacacactcagcctcgtatcatcttcaattgtaga 480
 DB 7636 CTTTAAAAAACTATACAAACACATTAACCTCAACTAATCATTTCTTCAATTATAA 7577
 OY 481 caagctgcgaagaagcttgagta 503
 DB 7576 CAAACTACCAAAAAAACTTAATA 7554

RESULT 3
 AAS46481
 ID AAS46481 standard; DNA; 12426 BP.

XX AAS46481;
 AC
 XX
 XX 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #203.
 XX
 XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 XX WO200168912-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 15-MAR-2001; 2001WO-EP02955.
 PF
 XX

PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A. Plegenbrock C, Berlin K.

DR WPI: 2001-602752/68.

PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer

PS Claim 1: SEQ ID No 203; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 12426 BP; 3527 A; 137 C; 2591 G; 6171 T; 0 other;

Query Match 67.8%; Score 343.8; DB 22; Length 12426;
 Best Local Similarity 79.9%; Pred. No. 4e-88;
 Matches 405; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

OY 1 tgaagtaatccatcacagagatttaaaagaactctctgcgcacccatcaacaata 60
 DB 4371 tgaagtaatccatcacagagatttaaaagaactctctgcgcacccatcaacaata 4430
 OY 61 tctccaccaatttgtagtaataatttgccaatgacccaatttttagtaggagcaat 120
 DB 4431 tttttatcaatttgtagtaataatttgccaatgacccaatttttagtaggagcaat 4490
 OY 121 ggtttaacataatcaatccctctctttttgattgcttcgtttaatgagcaacta 180
 DB 4491 ggtttaacataatcaatccctctcttttttttggattgcttcgtttaatgagcaacta 4550
 OY 181 cctctctggacatcagagatgagctgctgcagtttaacaaaaagaatgagatcag 240
 DB 4551 ttttttggatcttgtagtaagtagtctgcttatacaaaagaatgagatgag 4610
 OY 241 agtaattttgtgcacacagctgctcgaagaatttgtagtctactatcacacacatt 300
 DB 4611 agtaattttgtgcacacagctgctcgaagaatttgtagtctactatcacacacatt 4670
 OY 301 actttattcatcgaataattccaccctcgctcgtcgtagcgagagagattgcgta 360
 DB 4671 actttattcatcgaataattccaccctcgctcgtcgtagcgagagagattgcgta 4730

PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251866.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0259678.
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	

PI Rosen CA, Barash SC, Rubben SM,
XX
DR WPT: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure: SEQ ID NO 23775; 3071P + Sequence listing; English.

CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAM64703
CC to AAM87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAM54942 to AAM54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XQ
XQ Sequence 1239 BP; 391 A; 191 C; 176 G; 481 T; 0 other;

Query Match	7.58;	Score 38.2;	DB 22;	Length 1239;
Best Local Similarly	48.48;	Pred. No. 0.56;		
Matches 106; Conservative	0;	Mismatches 113;	Indels 0;	Gaps 0;

QY 1 tgaataatcccaacacacaggaattttaaagaagaactccgcgcacccattaacata 60
Db 82 ttagttatcatcaagaagacattgttaagtatacttcatccaagcttaataatc 141
QY 61 tctccaccaatttggtagtaaatatttgcataagtagtaatttttagtaggaact 120
Db 142 ttcttaataataatagacaacacttcttcacaagaagaattgaaatactcgatagtc 201
QY 121 gtgttaacacataatccactcctcttttggattgtctcttgtttaatggcagcta 180

Dd 202 ttcatltagaatcaaatglttagcctttcataattgttlttgttccttttaagaacatgagcagt 261

Dy 181 cctctcttgycatctagcagaatgagctgcgtcgactta 219
| | | ||| | | | |
Db 262 ttgtgaagtcatagaagaaggcgatlaagaagagtaatta 300 .

RESULT	5
AAK68964	
ID	AAK68964 standard; DNA; 13537 BP

AC AAK68964;

DT 06-NOV-2001 (first entry)

DE	Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:23776.
DE	Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:23776.

KW Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;

KM cytostatic; gene therapy; vaccine; metastasis; ds

OS Homo sapiens.

PN W0200157182-A2

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0189874.

PR 18-APR-2000; 2000US-0198123.

PR 07-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227009.

PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229344-
PT 01-SEP-2000; 2000US-0229345-PR 05-SEP-2000; 2000US-0229509.
DE-SEP-2000; 2000US-0338513

PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

XX	RESULT	6
XX	AAS46793	
ID	AAS46793 standard; DNA: 56153 BP.	
XX		
AC	AAS46793;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Tumour suppressor gene derived chemically modified sequence #519.	
XX		
KW	Human: tumour suppressor gene; oncogene; antitumour; cytostatic;	
KM	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;	
XX	cytosine methylation; ds.	
OS	Homo sapiens.	
XX		
PN	WO200168912-A2.	
PD		
XX	20-SEP-2001.	
PF		
XX	15-MAR-2001; 2001WO-EP02955.	
PR		
XX	15-MAR-2000; 2000DE-1013847.	
PR	06-APR-2000; 2000DE-1019058.	
PR	07-APR-2000; 2000DE-1019173.	
PR	30-JUN-2000; 2000DE-1032529.	
PR	01-SEP-2000; 2000DE-1043826.	
XX		
PA	(EPIC-) EPIGENOMICS AG.	
PI		
PI	Olek A, Piepenbrock C, Berlin K;	
DR	WPI: 2001-602752/68.	
XX		
PT	Fragments of chemically modified genes associated with tumour suppressor	
PT	genes and oncogenes, useful in designing primers and probes for	
PT	analysing diseases associated with cytosine methylation state e.g.	
PT	cancer	
XX		
PS	Claim 1; SEQ ID NO 519; 27pp; English.	
CC	The invention relates to a nucleic acid comprising a sequence of 18	
CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with	
CC	bisulphite, of genes associated with tumour suppression and	
CC	oncogenes having a sequence taken from 536 (actually 533 since	
CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences	
CC	(Ss) and sequences complementary to (Ss). The nucleic acid may be a	
CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may	
CC	form part of a set of probes for detecting the cytosine methylation state	
CC	and/or single nucleotide polymorphisms and also to be used in an	
CC	array for analysing diseases associated with CpG dinucleotides e.g.	
CC	cancers and tumours. The probes can also be used in a method for	
CC	ascertaining genetic and/or epigenetic parameters for the diagnosis	
CC	and/or therapy of existing diseases or the predisposition to specific	
CC	diseases, by analysing cytosine methylations. The parameters may be	
CC	compared to another set of genetic and/or epigenetic parameters, the	
CC	differences serving as basis for diagnosis and/or prognosis events which	
CC	are disadvantageous to patients. The present sequence is one of the	
CC	533 genomic sequences derived from tumour suppressor genes and	
CC	oncogenes.	
CC	Note: The sequence data for this patent did not form part	
CC	of the printed specification, but was obtained in electronic	
CC	format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pct_sequences.	
XX		
SQ	Sequence 56153 BP: 13943 A; 1002 C; 14095 G; 27113 T; 0 other:	
Query Match	7.5%; Score 38; DB 22; Length 56153;	
Best Local Similarity	47.5%; Pred No. 2.8; Indels	
Matches	113; Conservative 0; Mismatches 135; Gaps 0	
73	ttagtgagtaaatatttgcataagaccataatttagtagtcgacgtgtttatcacat	132

Db 31726 tctggagctataggtacgagttacttgctttaggtcttcgtcttcttgcattcttaattc 31785
Oy 133 ataccatccctctcttcttcttgaatgctcttcgtttaagtggcgacctcttgca 192
Db 31786 ttatgagttctttcttatatagtattggttaattcttcglttaatttttaatttgc 31845
Oy 193 tctagaagaatgagctgctgcagtttacacaagaatggaatcagagtaactttgtc 252
Db 31846 ttctcgtagtataaatatttatttaattctgtttgagaatgatctaatttagtcttaaaattc 31905
Oy 253 gccaccaagctgctcgtgaagaatttgttagtgctctatcacacacttaattattc 310
Db 31906 ttaggttatttttttaggtcttttttagttcttgcgaatgtaacttgcgttaagttttt 31963

RESULT 7
AAS46337
ID AAS46337 standard; DNA; 6063 BP.
XX
AC AAS46337;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #59.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KM cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KM cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI: 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
PS
PS Claim 1; SEQ ID NO 59; 27pp; English.

[illegible]

Query Match	Best Local Similarity	7.3%	Score 37;	DB 21;	Length 7143;
Matches 79;	Conservative 0;	Mismatches 70;	Indels 0;	Gaps 0	
Qy	70	aattggtgagtaaatattctgtctaatagtacctaatctttagtgtaggcagctgtttata	129		
Db	702	ATTTTTATATATATATATTTTATATATACATATATCTTAAGSGAGCATTTTTTTTATT	643		
Qy	130	catataccatctctctcttttttgatgctcttcctgtttaatgagcgagctactctctg	189		
Db	642	CATTATATCCATTATCATTTATTTTATTTATATATATCTTTATATATGTGAAGAAAAATA	583		
Qy	190	gcacttagcagaaagacgtcgtcagttt	218		
Db	582	GAAAGAAATMAAAATTAGCAATTTCTATT	554		
RESULT	9	AA161373/c			
ID	AA161373	standard; DNA; 513445 BP.			
XX	AA161373;				
XX	16-OCT-2001	(first entry)			
XX	Soybean	318013 region A3, SEQ ID NO: 4.			
XX	Soybean;	antihelminthic; gene therapy; soybean cyst nematode; SCN;			
XX	SCN	resistance; thgl; Rhg4; SCN resistant allele; plant breeding;			
XX	240017	region G3; 318013 region A3; 515002 region G2; ds.			
XX	Glycine	max.			
XX	MO200151627-A2.				
XX	19-JUL-2001.				
XX	05-JAN-2001;	2001WO-US000552.			
XX	07-JAN-2000;	2000US-0174880.			
XX	(MONS)	MONSANTO CO.			
XX	Hauge BM,	Wang ML, Parsons JD, Parnell LD;			
XX	WPI: 2001-425872/45.				
XX	P-PsDB; AAM42216.				

XX New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs -
XX
PS Claim 30: Page 596-893; 1353pp; English.
XX
XX The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule
CC provided in the specification.
XX
SQ Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;

Query Match
Best Local Similarity 7.3%; Score 36.8; DB 22; Length 513445;
Matches 130; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

OY 56 caatactccctacccaattgtagtaataatttgctaaagtaactaatttttagtag 115
DB 336067 CCATAGTCATCTACTTCTTACAAATTCCTCTCAATTTGGAAATACCTTTTATTTCT 336008
OY 116 gaactggtttacacatacatcctcctctttttagtgccttcgtttaagggc 175
DB 336007 GCTTAGAGTTTGGAAAAATATATCTAGTATGTTGATGATGGAATGAAGG 335948
OY 176 agctactctcttgtagatcag-aatgagctgctgcagtttaacaaaagaatga 234
DB 335947 ACTCATATCCCTGGTAAGAAATTAAGTTGATCAGTTTCAAGTATCAAAAAGAAATGT 335888
OY 235 gatcagagactcttctgcccacacagctcgtcgaagaattttagtctactatca 294
DB 335887 CACGAGAGATTAATCTGTTTATATATGTCGCGTAATGATTAAGTTCAATTTATCG 335828
OY 295 cacattacttattcatcagaatatt 322
DB 335827 CCCAGAAAAATTAATCGAAAAAATAAT 335800

RESULT 10
ABL3534
ID ABL3534 standard; DNA; 8964 BP.
XX
AC ABL3534;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1507.
XX
XX Human: immune system disease; cytosine methylation; antiasthmatic;
KM antileukosclerotic; antianaemic; cyrostatic; nootropic;
KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KM antileukemic; cancer; eye disease; arteriosclerosis; anaemia;
KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KM gene: ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX

PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Plepenbrock C, Berlin K;
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1: SEQ ID NO 1507; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 8964 BP; 2596 A; 89 C; 1815 G; 4464 T; 0 other;

Query Match
Best Local Similarity 7.2%; Score 36.6; DB 24; Length 8964;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 72 tttagtagtaaatatttgcataagtagtaactttagtagtagcagctgttatca 131
DB 7681 ttaattagaataatttcttgtaattatagaagaggttttggagaggttttttagattga 7740
OY 132 tatatcatcctctcttcttgtagtcttctgtttaatggcagctactctctggc 191
DB 7741 ttttttgcattagatttgcattattttagaggttctgttattgtttattat 7800
OY 192 atctagcagaatgagctgctgcagtt 218
DB 7801 ttgtagttgtagtgcgttttagtt 7827

RESULT 11
AAS61276
ID AAS61276 standard; DNA; 8964 BP.
XX
AC AAS61276;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #231.
XX
XX Human: Gene regulation-associated gene; severe combined immunodeficiency;
KM cardiac damage; inflammatory response; Hemophilia; Werner syndrome;
KM asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KM renal disease; Preeclampsia; cardiac allograft vascular disease;
KM colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KM immunostimulant; cardiac; anti-inflammatory; coagulant; antiasthmatic;
KM nephrotropic; gynecological; anti-tumour; immunosuppressive; cyrostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
XX

PR	29-SEP-2000	2000US-0236567	
PR	29-SEP-2000	2000US-0236568	
PR	29-SEP-2000	2000US-0236569	
PR	29-SEP-2000	2000US-0236570	
PR	02-OCT-2000	2000US-0236602	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	02-OCT-2000	2000US-0237039	
PR	13-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239335	
PR	13-OCT-2000	2000US-0239337	
PR	20-OCT-2000	2000US-0240960	
PR	20-OCT-2000	2000US-0241121	
PR	20-OCT-2000	2000US-0241185	
PR	20-OCT-2000	2000US-0241786	
PR	20-OCT-2000	2000US-0241807	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241809	
PR	01-NOV-2000	2000US-0241826	
PR	01-NOV-2000	2000US-0244617	
PR	08-NOV-2000	2000US-0246474	
PR	08-NOV-2000	2000US-0246475	
PR	08-NOV-2000	2000US-0246476	
PR	08-NOV-2000	2000US-0246477	
PR	08-NOV-2000	2000US-0246527	
PR	08-NOV-2000	2000US-0246532	
PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246619	
PR	08-NOV-2000	2000US-0246619	
PR	08-NOV-2000	2000US-0246613	
PR	17-NOV-2000	2000US-0249207	
PR	17-NOV-2000	2000US-0249208	
PR	17-NOV-2000	2000US-0249209	
PR	17-NOV-2000	2000US-0249210	
PR	17-NOV-2000	2000US-0249211	
PR	17-NOV-2000	2000US-0249212	
PR	17-NOV-2000	2000US-0249213	
PR	17-NOV-2000	2000US-0249214	
PR	17-NOV-2000	2000US-0249215	
PR	17-NOV-2000	2000US-0249216	
PR	17-NOV-2000	2000US-0249265	
PR	17-NOV-2000	2000US-0249297	
PR	17-NOV-2000	2000US-0249317	
PR	17-NOV-2000	2000US-0249444	
PR	17-NOV-2000	2000US-0249454	
PR	17-NOV-2000	2000US-0249464	
PR	17-NOV-2000	2000US-0249465	
PR	17-NOV-2000	2000US-0249466	
PR	17-NOV-2000	2000US-0249467	
PR	17-NOV-2000	2000US-0249468	
PR	17-NOV-2000	2000US-0249469	
PR	17-NOV-2000	2000US-0249470	
PR	17-NOV-2000	2000US-0249471	
PR	17-NOV-2000	2000US-0249472	
PR	17-NOV-2000	2000US-0249473	
PR	17-NOV-2000	2000US-0249474	
PR	17-NOV-2000	2000US-0249475	
PR	17-NOV-2000	2000US-0249476	
PR	17-NOV-2000	2000US-0249477	
PR	17-NOV-2000	2000US-0249478	
PR	17-NOV-2000	2000US-0249479	
PR	17-NOV-2000	2000US-0249480	
PR	17-NOV-2000	2000US-0249481	
PR	17-NOV-2000	2000US-0249482	
PR	17-NOV-2000	2000US-0249483	
PR	17-NOV-2000	2000US-0249484	
PR	17-NOV-2000	2000US-0249485	
PR	17-NOV-2000	2000US-0249486	
PR	17-NOV-2000	2000US-0249487	
PR	17-NOV-2000	2000US-0249488	
PR	17-NOV-2000	2000US-0249489	
PR	17-NOV-2000	2000US-0249490	
PR	17-NOV-2000	2000US-0249491	
PR	17-NOV-2000	2000US-0249492	
PR	17-NOV-2000	2000US-0249493	
PR	17-NOV-2000	2000US-0249494	
PR	17-NOV-2000	2000US-0249495	
PR	17-NOV-2000	2000US-0249496	
PR	17-NOV-2000	2000US-0249497	
PR	17-NOV-2000	2000US-0249498	
PR	17-NOV-2000	2000US-0249499	
PR	17-NOV-2000	2000US-0249500	
PR	17-NOV-2000	2000US-0249501	
PR	17-NOV-2000	2000US-0249502	
PR	17-NOV-2000	2000US-0249503	
PR	17-NOV-2000	2000US-0249504	
PR	17-NOV-2000		

Query Match	7.2%	Score 36.4	DB 22	Length 44211
Best Local Similarity	56.8%	Pred. No. 7.4		
Matches 67	Conservative 0	Mismatches 51	Indels 0	Gaps 0
Oy	80	taaatatttgcataagtagtaacctaatctttagtgtaggcacgtgtcttatacatcca	139	
Db	11156	TAAAAAGGTTTCACGACAGATTGAATGTTATTATTATATATATATATATATATATA	11097	
Oy	140	ttctctcttcttggcttgccttcgttcaatcggcagctacacctcttgcatctag	197	
Db	11096	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGCGACAGGTCCTGTTGTGCCACG	11039	
RESULT 13				
ID	AA546787/C			
XX	AA546787 standard; DNA; 61020 BP.			
XX	AA546787;			
XX				
XX	18-DEC-2001 (first entry)			
XX				
XX	Tumour suppressor gene derived chemically modified sequence #513.			
XX				
XX	Human; tumour suppressor gene; oncogene; antitumour; cytostatic;			
XX	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;			
XX	cytosine methylation; ds.			
XX				
XX	Homo sapiens.			
XX				
XX	WO200168912-A2.			
XX				
XX	20-SEP-2001.			
XX				
XX	15-MAR-2001; 2001WO-EP02955.			
XX				
XX	15-MAR-2000; 2000DE-1013847.			
XX	06-APR-2000; 2000DE-1019058.			
XX	07-APR-2000; 2000DE-1019173.			
XX	30-JUN-2000; 2000DE-1032529.			
XX	01-SEP-2000; 2000DE-1043826.			
XX				
XX	(EPIG-) EPIGENOMICS AG.			
XX				
XX	Olek A. Pfenberock C. Berlin K;			
XX				

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 14:45:19 ; Search time 3309.28 Seconds
(without alignments)
2067.808 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507

Sequence: 1 tgagtaatcctaatacagg.....ccaagaggcttgtagtaggag 507

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.4	8.6	892	12 A2539346	A2539346 ENTGP83TR
2	41.8	8.2	1201	12 CNS016FX	AL106695 Drosophil
3	40.6	8.0	681	12 A0421326	A0421326 RPCI-11-1
4	40.2	7.9	1042	12 CNS024NB	AL180992 Tetraodon
5	40	7.9	1006	12 CNS03ESG	AL240793 Tetraodon
6	39.8	7.9	774	9 AL546807	AL546807 AL546807
7	39.6	7.8	457	10 BE943660	BE943660 UI-M-BH3-
8	39.6	7.8	494	9 A1848569	A1848569 UI-M-AM1-
9	39.6	7.8	1142	10 BM464476	BM464476 AGENCOURT
10	39.2	7.7	928	12 CNS00DKY	AL071865 Drosophil
11	39	7.7	850	12 A2673135	A2673135 ENTLI30TR
12	38.8	7.7	648	12 BH587629	BH587629 BOHEX88TF
13	38.6	7.6	365	9 A1605224	A1605224 ve78a05.x
14	38.6	7.6	388	12 B58764	B58764 CIT-RSP-201
15	38.6	7.6	419	10 BM163067	BM163067 EST565590
16	38.6	7.6	711	10 BF468105	BF468105 UI-M-CD0-
17	38.4	7.6	443	12 A0450690	A0450690 HS_5161_A

c 18	38.4	7.6	507	12 CNS01852	AL108896 Drosophil
19	38.4	7.6	726	10 BM170969	BM170969 EST573492
c 20	38.4	7.6	884	12 CNS00CXG	AL060063 Drosophil
c 21	38.4	7.6	1099	12 CNS01481	AL103836 Drosophil
22	38.4	7.6	1101	12 CNS0039G	AL063921 Drosophil
23	38.2	7.5	641	12 AG066397	AG066397 Pan trogl
24	38.2	7.5	907	12 CNS07COV	AL439445 T7 end of
25	37.8	7.5	535	12 AZ748844	AZ748844 RPCI-24-1
c 26	37.8	7.5	878	12 CNS0187R	AL108993 Drosophil
c 27	37.8	7.5	932	12 CNS0180Y	AL108748 Drosophil
c 28	37.8	7.5	935	12 CNS00J34	AL076343 Drosophil
c 29	37.8	7.5	1023	12 CNS00CVF	AL059889 Drosophil
c 30	37.8	7.5	1101	12 CNS008X3	AL052544 Drosophil
c 31	37.6	7.4	312	12 AZ821437	AZ821437 2M0094009
c 32	37.6	7.4	396	12 AQ683613	AQ683613 HS_5458_B
33	37.6	7.4	873	12 CNS024M5	AL180950 Tetraodon
c 34	37.4	7.4	797	10 BF253656	BF253656 HVSME4000
c 35	37.4	7.4	1007	12 CNS01413	AL104181 Drosophil
c 36	37.4	7.4	1101	12 CNS0006J	AL062049 Drosophil
c 37	37.2	7.3	353	9 AW627276	AW627276 pa28a06.y
c 38	37.2	7.3	589	10 BE886609	BE886609 601509355
c 39	37.2	7.3	607	12 AG068424	AG068424 Pan trogl
c 40	37.2	7.3	614	9 AL655660	AL655660 AL655660
c 41	37.2	7.3	761	12 BH470789	BH470789 BOGYM20TR
c 42	37.2	7.3	767	12 CNS00AQX	AL055924 Drosophil
43	37.2	7.3	865	12 AZ675018	AZ675018 ENTHS57TR
44	37.2	7.3	889	12 AZ543817	AZ543817 ENFTS3TF
c 45	37	7.3	500	9 AU087167	AU087167 AU087167

ALIGNMENTS

RESULT 1

AZ539346 892 bp DNA linear GSS 14-NOV-2000
LOCUS ENTGP83TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION A2539346
VERSION A2539346.1 GI:11145135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 892)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 847.

FEATURES

Location/Qualifiers
1..892
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS01; Site:1; Bst.I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

Qy	78	agtaaatatttgcctaatagtacctaatttttaggtaggcactgcttatcatatc	139
		- -	
Db	386	AATAAATCTATTCTAAATAATAATATATTTTATTAATAATATTATATACCTTTATT	415

三


```

Query Match      7.8%; Score 39.6; DB 10; Length 457;
Best Local Similarity 66.3%; Pred. No. 41;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaatttctgaatagtagtaattttagtaggcactgtgtttataca 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 TTTCGAAGTCAATATTTAGCTGATGTAATTAATATTTCTTAGGAATGTTGGTTATCCA 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 tatatcattccctcttttttgattg 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 ACTACCTATTTTCTTTTACTTTG 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
LOCUS      A1848569          494 bp      mRNA      linear      EST 15-JUL-1999
DEFINITION UI-M-AM1-aga-b-02-0-UI.s2 NIH_BMAP_MAM_N Mus musculus cDNA clone
ACCESSION  A1848569
VERSION    A1848569.1 GI:5492475
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus

REFERENCE  1
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Contact: Chin, H
          National Institute of Mental Health
          6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
          20892-9643, USA
          Tel: 301 443 1706
          Fax: 301 443 9890
          Email: mEST@mail.nih.gov
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. The sequence tag present in the cDNA between the NotI site
          and the oligo-dT track served to verify it as a clone from the
          normalized amygdala library cDNA Library Preparation: M.B. Soares
          Lab Clone distribution: NIH BMAP cDNA clones will be made available
          by the means that is soon to be determined. When NIH determines the
          means for distribution of the BMAP cDNA clones, this record will be
          updated accordingly when that means is determined.
          Seq primer: M13 Forward
          POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..494
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AM1-aga-b-02-0-UI"
                     /clone_lib="NIH_BMAP_MAM_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker. Site 1: Not I; Site 2: Eco RI; The
                     NIH_BMAP_MAM_N library is a normalized library constructed
                     from mouse amygdala. The tag is a string of 5 nucleotides
                     present between the Not I site and the oligo-dT track.
                     The library was constructed as described by Bonaldo,
                     Lennon and Soares, Genome Research 6: 791-806, 1996.
                     Tissue provided by Ms. Annie Novakovich, Zivic-Miller
                     Laboratories.
                     TAG_LIB=NIH_BMAP_MAM_N
                     TAG_TISSUE=amygdala
                     TAG_SEQ=GTGAG"

BASE COUNT      126 a 118 c 91 g 159 t

```

ORIGIN

```

Query Match      7.8%; Score 39.6; DB 9; Length 494;
Best Local Similarity 56.3%; Pred. No. 41;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaatttctgaatagtagtaattttagtaggcactgtgtttataca 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 TTTCGAAGTCAATATTTAGCTGATGTAATTAATATTTCTTAGGAATGTTGGTTATCCA 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 tatatcattccctcttttttgattg 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 ACTACCTATTTTCTTTTACTTTG 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS      BM464476          1142 bp      mRNA      linear      EST 05-FEB-2002
DEFINITION AGENCOURT_6438833 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535563
ACCESSION  BM464476
VERSION    BM464476.1 GI:18513518
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

REFERENCE  1 (bases 1 to 1142)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12223 row: n column: 12
          High quality sequence stop: 630.

FEATURES             Location/Qualifiers
     source           1..1142
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5535563"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb."
                     /seq_id_start=378 a 245 c 224 g 259 t 36 others

BASE COUNT      378 a 245 c 224 g 259 t 36 others
ORIGIN

Query Match      7.8%; Score 39.6; DB 10; Length 1142;
Best Local Similarity 48.6%; Pred. No. 39;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 105 tttttagtaggcactgttttatcatatcatcctctttttttttttttttttct 164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 TTTTNNNNNNNTTTTNTNTTTTNTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 gtttaattggcagctctctcttggcctatgcagaaagcgtcgtcagtttacacaa 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 TTTTTCGCACATATACCACTTACTTTATGTACAAATAAGGAATGGGAGGGGAAATC 526
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 225 aaagaatggagatcagagtcactttttgtgccaccaacgtgtctgagaaatttgtt 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 AAAGAATAGAGAAACTATACGGTAGCTAGCTAGGATGTGTGTGGAACCAAAATTCAGTTT 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:52:52 ; Search time 3102.12 seconds
(without alignments)
296.819 Million cell updates/sec

Title: US-09-617-174B-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagaggcttagtaggag 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
-----	-----	-----	-----	-----

1	42.4	96.4	170269	9	AC090307	Homo sapi
2	42.4	96.4	173131	2	AC036176	Homo sapi
3	42.4	96.4	173836	2	AP001404	Homo sapi
4	42.4	96.4	214212	2	AC015793	Homo sapi
5	42.4	96.4	214212	2	AC015793	Homo sapi
6	31.2	70.9	12426	6	AX251235	Sequence
7	27.6	62.7	177842	2	AC025047	Mus muscu
8	24.8	56.4	117032	9	AC008973	Homo sapi
9	24.8	56.4	139476	9	AC008773	Homo sapi
10	24.8	56.4	141557	2	AC008455	Homo sapi
11	24.8	56.4	155459	2	AC016619	Homo sapi
12	24.8	56.4	174831	9	AC008966	Homo sapi
13	24	54.5	90309	2	AC098644	Takifugu
14	24	54.5	157241	2	AC046194	Homo sapi
15	24	54.5	194612	9	AC021118	Homo sapi
16	23.8	54.1	47872	7	AF338467	Cyanophag
17	23.6	53.6	61922	2	AC109216	Mus muscu
18	23.4	53.2	104712	9	AC079412	Homo sapi
19	23.4	53.2	149180	2	AC015525	Homo sapi
20	23.4	53.2	183099	2	AC025288	Homo sapi
21	23.4	53.2	205978	2	AC007491	Homo sapi
22	23.2	52.7	708	33	AC056333	Giardia
23	23.2	52.7	877	33	AC029184	Giardia
24	23.2	52.7	914	33	AC056332	Giardia
25	23.2	52.7	928	33	AC055027	Giardia
26	23.2	52.7	974	33	AC034502	Giardia
27	23.2	52.7	1057	33	AC086750	Giardia
28	23.2	52.7	135684	9	AC002463	Human BAC
29	23.2	52.7	136344	2	AC106410	Rattus no
30	23.2	52.7	162714	2	AC027569	Homo sapi
31	23.2	52.7	168637	9	AC073137	Homo sapi
32	23.2	52.7	173975	2	AC073045	Homo sapi
33	23.2	52.7	179389	9	AC012558	Homo sapi
34	23.2	52.7	197321	9	AF315553	Homo sapi
35	23.2	52.7	203665	2	AC092179	Homo sapi
36	23.2	52.7	208307	2	AC073351	Mus muscu
37	23.2	52.7	214587	2	AC074332	Mus muscu
38	23.2	52.7	222804	9	AC018977	Homo sapi
39	23	52.3	1820	4	PIGSPABIND	L10124 Sus scrofa
40	23	52.3	13120	1	AE009939	Pyrobacul
41	22.8	51.8	924	3	CDU21211	Chironomus
42	22.8	51.8	2039	3	CHIHH11	Chironomus
43	22.8	51.8	2140	3	CHIHH1F	Chironomus
44	22.8	51.8	2375	9	AF071476	Homo sapi
45	22.8	51.8	6696	3	CTHIS	C.thummi ge

ALIGNMENTS

RESULT 1	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
AC090307	AC090307	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.			
LOCUS	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
DEFINITION	AC090307	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.			
ACCESSION	AC090307				
VERSION	AC090307.7	GI:16041402			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1	(bases 1 to 170269)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.				
TITLE	Homo sapiens chromosome 18, clone RP11-851B10				
JOURNAL	Unpublished				
REFERENCE	2	(bases 1 to 170269)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgeert, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,				

Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldrim, J., Meneus, L., Mihoval, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 170269)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bustien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
 Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihoval, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL
 Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Oct 11, 2001 this sequence version replaced gi:15290867.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L12576
 Center clone name: 851_B_10

FEATURES

source

Location/Qualifiers
 1. 170269

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-851B10"
 /clone_lib="RPC1-11 Human Male BAC"
 complement(2..177)
 /rpt_family="L1LNC5"
 complement(189..489)
 /rpt_family="AluSx"
 complement(708..803)
 /rpt_family="MIR"
 complement(1616..1904)
 /rpt_family="AluJo"
 complement(2815..3120)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/rpt_family="AluSx"
 3218..3365
 /rpt_family="MIR"
 3565..3590
 /rpt_family="(TC)n"
 complement(3782..4028)
 /rpt_family="MIR"
 4056..4177
 /rpt_family="MER5A"
 4601..5620
 /rpt_family="HSMAR2"
 5621..5643
 /rpt_family="(CAA)n"
 5644..5866
 /rpt_family="HSMAR2"
 6184..6475
 /rpt_family="L1MA2"
 6677..6837
 /rpt_family="(TA)n"
 6852..7018
 /rpt_family="(TA)n"
 6994..7122
 /rpt_family="(CATATA)n"
 7146..7332
 /rpt_family="(CATATA)n"
 7371..7526
 /rpt_family="(TA)n"
 complement(7527..7835)
 /rpt_family="AluSx"
 7936..8080
 /rpt_family="MIR"
 8697..8927
 /rpt_family="AluSx"
 complement(9036..9288)
 /rpt_family="AluSg"
 9873..9934
 /rpt_family="(GA)n"
 10340..10367
 /rpt_family="AT-rich"
 10368..10686
 /rpt_family="AluY"
 10687..10711
 /rpt_family="AT-rich"
 complement(10870..11043)
 /rpt_family="L2"
 11348..11375
 /rpt_family="(T)n"
 12359..12449
 /rpt_family="L3"
 complement(12634..12890)
 /rpt_family="MIR"
 complement(13436..13725)
 /rpt_family="AluJb"
 14244..14297
 /rpt_family="(CGGG)n"
 complement(15188..15350)
 /rpt_family="MIR3"
 16056..16114
 /rpt_family="MIR3"
 complement(16129..16429)
 /rpt_family="AluSx"
 16532..16646
 /rpt_family="MER94"
 16855..17145
 /rpt_family="L2"
 17377..17572
 /rpt_family="MIR3"
 complement(20036..20330)
 /rpt_family="Charliel"
 complement(20333..20486)
 /rpt_family="Charliel"
 20625..20691
 /rpt_family="L2"

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 13, 2001 this sequence version replaced gl:16949388.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L9215
Center clone name: 635_N_19

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 114304: contig of 114304 bp in length
* 114305 114404: gap of 100 bp
* 114405 173131: contig of 58727 bp in length.

Location/Qualifiers
1. .173131
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-635N19"
/clone_lib="RPC1-11 Human Male BAC"

BASE COUNT 51584 a 35315 c 35660 g 50470 t 101 others

ORIGIN

Query Match 95.4%; Score 42.4; DB 2; Length 173131;
Best Local Similarity 97.7%; Pred. NO. 1.2e-07;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tctcttcaattgtgacaagctgccaaagagcgttgagtaggag 44
+|||||
Db 51903 TTCTTTCATTTGGACAAGCTGCCAAGAGCGCTTGAGTAGGAG 51860

RESULT 3
AP001404/c

LOCUS AP001404 Homo sapiens chromosome 18 clone RP11-651B10 map 18q22, WORKING
DEFINITION DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AP001404.3 GI:5188495
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens DNA, clone:RP11-851B10.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173836)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 173,836 genomic DNA of 18q22

TITLE

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE JOURNAL

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 3, 2000 this sequence version replaced gi:6446894.

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: L117_P_15

* NOTE: This record contains 240 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 1 695: contig of 695 bp in length
 * 696 795: gap of 100 bp
 * 796 1513: contig of 718 bp in length
 * 1514 1613: gap of 100 bp
 * 1614 2337: contig of 724 bp in length
 * 2338 2437: gap of 100 bp
 * 2438 3161: contig of 724 bp in length
 * 3162 3261: gap of 100 bp
 * 3262 3966: contig of 705 bp in length
 * 3967 4066: gap of 100 bp
 * 4067 4772: contig of 706 bp in length
 * 4773 4872: gap of 100 bp
 * 4873 5564: contig of 692 bp in length
 * 5565 5664: gap of 100 bp
 * 5665 6397: contig of 733 bp in length
 * 6398 6497: gap of 100 bp
 * 6498 7202: contig of 705 bp in length
 * 7203 7302: gap of 100 bp
 * 7303 8028: contig of 726 bp in length
 * 8029 8128: gap of 100 bp
 * 8129 8803: contig of 675 bp in length
 * 8804 8903: gap of 100 bp
 * 8904 9613: contig of 710 bp in length
 * 9614 9713: gap of 100 bp
 * 9714 10409: contig of 696 bp in length
 * 10410 10509: gap of 100 bp
 * 10510 11237: contig of 728 bp in length
 * 11238 11337: gap of 100 bp
 * 11338 12043: contig of 706 bp in length
 * 12044 12143: gap of 100 bp
 * 12144 12867: contig of 724 bp in length
 * 12868 12967: gap of 100 bp
 * 12968 13660: contig of 693 bp in length
 * 13661 13760: gap of 100 bp
 * 13761 14489: contig of 729 bp in length
 * 14490 14589: gap of 100 bp
 * 14590 15314: contig of 725 bp in length
 * 15315 15414: gap of 100 bp
 * 15415 16129: contig of 715 bp in length
 * 16130 16229: gap of 100 bp
 * 16230 16946: contig of 717 bp in length
 * 16947 17046: gap of 100 bp
 * 17047 17769: contig of 723 bp in length

* 17770 17869: gap of 100 bp
 * 17870 18574: contig of 705 bp in length
 * 18575 18674: gap of 100 bp
 * 18675 19379: contig of 705 bp in length
 * 19380 19479: gap of 100 bp
 * 19480 20186: contig of 707 bp in length
 * 20187 20236: gap of 100 bp
 * 20287 20980: contig of 694 bp in length
 * 20981 21080: gap of 100 bp
 * 21081 21779: contig of 699 bp in length
 * 21780 21879: gap of 100 bp
 * 21880 22594: contig of 715 bp in length
 * 22595 22694: gap of 100 bp
 * 22695 23399: contig of 705 bp in length
 * 23400 23499: gap of 100 bp
 * 23500 24217: contig of 718 bp in length
 * 24218 24317: gap of 100 bp
 * 24318 25026: contig of 709 bp in length
 * 25027 25126: gap of 100 bp
 * 25127 25836: contig of 710 bp in length
 * 25837 25936: gap of 100 bp
 * 25937 26637: contig of 701 bp in length
 * 26638 26737: gap of 100 bp
 * 26738 27446: contig of 709 bp in length
 * 27447 27546: gap of 100 bp
 * 27547 28274: contig of 728 bp in length
 * 28275 28374: gap of 100 bp
 * 28375 29095: contig of 721 bp in length
 * 29096 29155: gap of 100 bp
 * 29156 29917: contig of 722 bp in length
 * 29918 30017: gap of 100 bp
 * 30018 30724: contig of 707 bp in length
 * 30725 30824: gap of 100 bp
 * 30825 31529: contig of 705 bp in length
 * 31530 31629: gap of 100 bp
 * 31630 32324: contig of 695 bp in length
 * 32325 32424: gap of 100 bp
 * 32425 33116: contig of 692 bp in length
 * 33117 33216: gap of 100 bp
 * 33217 33936: contig of 720 bp in length
 * 33937 34036: gap of 100 bp
 * 34037 34750: contig of 714 bp in length
 * 34751 34850: gap of 100 bp
 * 34851 35578: contig of 728 bp in length
 * 35579 35678: gap of 100 bp
 * 35679 36393: contig of 715 bp in length
 * 36394 36493: gap of 100 bp
 * 36494 37203: contig of 710 bp in length
 * 37204 37303: gap of 100 bp
 * 37304 37965: contig of 662 bp in length
 * 37966 38065: gap of 100 bp
 * 38066 33781: contig of 716 bp in length
 * 38782 38881: gap of 100 bp
 * 38882 39590: contig of 709 bp in length
 * 39591 39690: gap of 100 bp
 * 39691 40403: contig of 713 bp in length
 * 40404 40503: gap of 100 bp
 * 40504 41238: contig of 735 bp in length
 * 41239 41338: gap of 100 bp
 * 41339 42064: contig of 726 bp in length
 * 42065 42164: gap of 100 bp
 * 42165 42880: contig of 716 bp in length
 * 42881 42980: gap of 100 bp
 * 42981 43687: contig of 707 bp in length
 * 43688 43787: gap of 100 bp
 * 43788 44504: contig of 717 bp in length
 * 44505 44604: gap of 100 bp
 * 44605 45307: contig of 703 bp in length
 * 45308 45407: gap of 100 bp
 * 45408 46106: contig of 699 bp in length
 * 46107 46206: gap of 100 bp
 * 46207 46909: contig of 703 bp in length
 * 46910 47009: gap of 100 bp

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L1590
Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 695: contig of 695 bp in length
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12967 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14589: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp

Query Match 96.4%; Score 42.4; DB 2; Length 214212;
Best Local Similarity 97.7%; Pred. No. 1.2e-07;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 1 tctcttcaatttggaacagctgccaaagaggttgtagtaggag 44
Db 57083 TTTCTTTCAATTGTGGACAGCTCCCAAGAGGCTTGAGTAGGAG 57126

RESULT 5
AC015793/c
LOCUS AC015793
DEFINITION Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 214212)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 18, clone RP11-1117D15
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 214212)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6446894.

KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 177842)
JOURNAL Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE Mus musculus chromosome 4, clone RP23-5618
AUTHORS Unpublished
2 (bases 1 to 177842)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavsky,L., Bouknight,A., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choebel,Y., Collangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galegan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Weldrin,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 1, 2002 this sequence version replaced gi:18182780.
All repeats were identified using RepeatMasker:
Smit,A.F.P.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1603
Center clone name: 56_I-8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 61132: contig of 61132 bp in length
* 61133 61232: gap of 100 bp
* 61233 154046: contig of 92814 bp in length
* 154047 154146: gap of 100 bp
* 154147 177842: contig of 23696 bp in length.
FEATURES
Location/Qualifiers
1..177842
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/map="4"
/clone="RP23-5618"
/clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT 51914 a 37479 c 36380 g 51862 t 207 Others
ORIGIN
Query Match 62.7%; Score 27.6; DB 2; Length 177842;
Best Local Similarity 56.4%; Score 24.8; DB 9; Length 117032;
Matches 33; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
Qy 1 tctcttcaattgtgacaaagctgccaagaggttgcagtagg 42
Db 21723 TGCTCTCAATTTGGACACACGCGCCAGCTGAGTGAAG 21682
RESULT 8
AC008973/c 117032 bp DNA linear PRI 06-SEP-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2380L12, complete sequence.
DEFINITION AC008973
ACCESSION AC008973
VERSION AC008973.6 GI:15451655
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 117032)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 117032)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 117032)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 117032)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Sep 6, 2001 this sequence version replaced gi:10954347.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
FEATURES
Location/Qualifiers
1..117032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2380L12"
BASE COUNT 37142 a 22654 c 22476 g 34750 t
ORIGIN
Query Match 56.4%; Score 24.8; DB 9; Length 117032;
Best Local Similarity 72.7%; Pred. No. 6.4;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 1 tctcttcaattgtgacaaagctgccaagaggttgcagtagg 44
Db 20876 TCTCTCAATTTGGACACACGCGCCAGCTGAGTGAAG 20833
RESULT 9
AC008773/c 139476 bp DNA linear PRI 03-OCT-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2017119, complete sequence.
DEFINITION AC008773
ACCESSION AC008773
VERSION AC008773.8 GI:15887254
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 2 (bases 1 to 139476)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 3 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 4 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
COMMENT On Oct 3, 2001 this sequence version replaced gi:10312245.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.8.
STS Content:
SHGC-7426 GI4255
SHGC-36388 G30154
SHGC-16052 G15353.
FEATURES
source
1. 139476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-201719"
BASE COUNT 42042 a 26255 c 27397 g 43782 t
ORIGIN
Query Match 56.4%; Score 24.8; DB 9; Length 139476;
Best Local Similarity 72.7%; Pred. No. 6.5;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 tctcttcaattgtgacaaagctgccaaagagcttgagtaggag 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122755 TCTCTCTCAATTTTGGACACACGCGCCAGTGGACTTTAGTTGAAG 122712

RESULT 10
AC008455 141597 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTC-34711, WORKING DRAFT SEQUENCE,
DEFINITION 9 ordered pieces.
ACCESSION AC008455
VERSION AC008455.7 GI:13676949
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 141597)
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
UNPUBLISHED
REFERENCE
AUTHORS 2 (bases 1 to 141597)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
UNPUBLISHED
COMMENT On Apr 19, 2001 this sequence version replaced gi:10305080.

```

```

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 318619
Center clone name: CIT-HSPC_34711
-----
Summary Statistics
Consensus quality: 128407 bases at least Q40
Consensus quality: 138430 bases at least Q30
Consensus quality: 139955 bases at least Q20
Estimated insert size: 140000; pulse field gel estimation
Estimated insert size: 140797; sum-of-contigs estimation
Quality coverage: 5.6 in Q20 bases; pulse field gel estimation
Quality coverage: 5.57 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 26634: contig of 26634 bp in length
* 26635 26734: gap of unknown length
* 26735 34856: contig of 8122 bp in length
* 34857 34956: gap of unknown length
* 34957 61826: contig of 26870 bp in length
* 61827 61926: gap of unknown length
* 61927 65313: contig of 3387 bp in length
* 65314 65413: gap of unknown length
* 65414 67577: contig of 2164 bp in length
* 67578 67677: gap of unknown length
* 67678 100223: contig of 32546 bp in length
* 100224 100323: gap of unknown length
* 100324 125153: contig of 24830 bp in length
* 125154 125253: gap of unknown length
* 125254 133618: contig of 8365 bp in length
* 133619 133718: gap of unknown length
* 133719 141597: contig of 7879 bp in length.
FEATURES
Location/Qualifiers
1. 141597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-34711"
/clone_lib="CalTech human BAC library C"
BASE COUNT 41447 a 26594 c 26971 g 45784 t 801 others
ORIGIN
Query Match 56.4%; Score 24.8; DB 2; Length 141597;
Best Local Similarity 72.7%; Pred. No. 6.5;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 tctcttcaattgtgacaaagctgccaaagagcttgagtaggag 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135935 TCTCTCTCAATTTTGGACACACGCGCCAGTGGACTTTAGTTGAAG 135978

RESULT 11
AC016619/c 155459 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2245F16, WORKING DRAFT
DEFINITION SEQUENCE, 6 ordered pieces.
ACCESSION AC016619
VERSION AC016619.5 GI:7711578
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

```


Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 90309)
Green, E.D.
Direct Submission
Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
On Nov 30, 2001 this sequence version replaced gi:16506397.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@hgrl.nih.gov
----- Project Information
Center project name: arp
Center clone name: 205P11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 87445 bases at least Q40
Consensus quality: 88132 bases at least Q30
Consensus quality: 88439 bases at least Q20
Insert size: 89000; agarose-fp
Insert size: 89609; sum-of-contigs
Quality coverage: 10.67x in Q20 bases; agarose-fp
Quality coverage: 10.60x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2330: contig of 2330 bp in length
* 2331 2430: gap of unknown length
* 2431 5020: contig of 2590 bp in length
* 5021 5120: gap of unknown length
* 5121 7685: contig of 2565 bp in length
* 7686 7785: gap of unknown length
* 7786 13195: contig of 5410 bp in length
* 13196 13295: gap of unknown length
* 13296 20253: contig of 6958 bp in length
* 20254 20353: gap of unknown length
* 20354 26321: contig of 5968 bp in length
* 26322 26421: gap of unknown length
* 26422 51911: contig of 25490 bp in length
* 51912 52011: gap of unknown length in length.
* 52012 90309: contig of 38298 bp in length.
Location/Qualifiers
1..90309
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="205P11"
/clone_lib="Incyte Genomics"
1..2330
/note="assembly_fragment"
2431..5020
/note="assembly_fragment"
5121..7685
/note="assembly_fragment"
7786..13195
/note="assembly_fragment"
13296..20253
/note="assembly_fragment"
20354..26321
/note="assembly_fragment"
26422..51911

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

/note="assembly_fragment"
clone_end:T7
vector_side:left
52012..90309
/note="assembly_fragment"
clone_end:SP6
vector_side:right
23869 a 20700 c 20314 g 24692 t 734 others
BASE COUNT
ORIGIN

Query Match 54.5%; Score 24; DB 2; Length 90309;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 4 cttcaattgtgacaaactgccaagagcttgtagtagga 43
Db 32119 CCTTCTGTATGGACCACTGCCTACAGCTAAAGCAGGA 32158

RESULT 14

AC046194 AC046194 157241 bp DNA linear HTG 23-SEP-2000
LOCUS Homo sapiens chromosome 4 clone RP11-578A19 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 13 unordered pieces.
AC046194
VERSION AC046194.3 GI:10280914
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157241)
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 4, clone RP11-578A19
Unpublished
REFERENCE
2 (bases 1 to 157241)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choe, P., Dea, P., Dea, P., Collins, S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galligan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Plier, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,
Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8099902.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9619
Center clone name: 578_A_19

FEATURES
source

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

Db 91972 TTTCCATTTGGACGAGTTTTTTGATAGGCTTCAGTGGAG 92011
||||| ||| ||||| || | | ||||| ||||| |||||

Search completed: September 24, 2002, 15:53:42
Job time: 20317 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:26 : Search time 403.26 Seconds
(without alignments)
187.334 Million cell updates/sec

Title: US-09-617-174B-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagaggcttgagtaggag 44

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description		
1	44	100.0	1141	18	AAT89000		
c	31.2	70.9	12426	22	AA546481		
	23	52.3	1820	16	AAQ88817		
	21.4	66.15	23	ABL26203	Drosophila melanog		
	22.6	51.4	10565	23	ABL26202	Drosophila melanog	
	22.4	50.9	321	22	AAK50410	Human bone marrow	
c	7	22.4	50.9	349	22	AAK54915	Human haematologic
c	8	22.4	50.9	527	22	AAK37539	Human bone marrow
	9	22.4	50.9	12426	22	AA546482	Tumour suppressor

c	10	22.2	50.5	2280	23	ABL26459	Drosophila melanog
	11	22.2	50.5	4530	23	ABL26458	Drosophila melanog
	12	22.2	50.5	8642	23	ABL02384	Drosophila melanog
	13	22.2	50.5	74586	24	AAS16905	Genomic DNA encodi
	14	22	50.0	1318	23	AAS71982	DNA encoding novel
	15	21.6	49.1	500	22	ABA58422	Human foetal liver
	16	21.6	49.1	500	22	ABA27523	Probe #5989 for ge
	17	21.6	49.1	500	22	AAK06526	Human brain expres
	18	21.6	49.1	500	22	AAK32211	Human bone marrow
	19	21.6	49.1	500	22	AAI16008	Probe #5941 for ge
	20	21.6	49.1	500	22	AAI38062	Probe #6748 used t
	21	21.6	49.1	5585	21	AAA60446	Murine factor v en
	22	21.4	48.6	1422	22	AAI03825	Human reproductive
	23	21.4	48.6	1596	22	AAH34605	Human colon cancer
	24	21.4	48.6	1682	22	AAK71694	Human immune/haema
	25	21.4	48.6	1682	22	AAK71695	Human immune/haema
	26	21.4	48.6	1682	22	AAK71697	Human immune/haema
	27	21.4	48.6	2622	22	AAH18084	Human cDNA sequenc
	28	21	47.7	34279	22	AAF28539	Genomic fragment #
	29	21	47.7	140036	24	AAS98600	Human genomic DNA
	30	20.8	47.3	233	21	AAC21576	Human secreted pro
	31	20.8	47.3	458	22	AAK05085	Human brain expres
	32	20.8	47.3	458	22	AAK30641	Human bone marrow
	33	20.8	47.3	458	22	AAI36578	Probe #5264 used t
	34	20.8	47.3	966	24	ABI99550	Mouse ischaemic co
	35	20.8	47.3	2468	22	AAI59786	Human polynucleoti
	36	20.8	47.3	2868	22	AAH15964	Human cDNA sequenc
	37	20.8	47.3	4253	18	AAH6172	Human MLN 51 nucle
	38	20.8	47.3	4253	22	AAI58000	Human polynucleoti
	39	20.8	47.3	8396	21	AAZ59353	Human STP2 (phenol
	40	20.6	46.8	1038	23	ABL13843	Drosophila melanog
	41	20.6	46.8	1254	15	AAQ55782	Segment of human g
	42	20.6	46.8	1254	15	AAQ58718	Human gamma-1 chai
	43	20.6	46.8	2252	23	ABL14548	Drosophila melanog
	44	20.6	46.8	2705	23	ABL05757	Drosophila melanog
	45	20.6	46.8	2949	23	ABL22984	Drosophila melanog

ALIGNMENTS

RESULT 1

AAT89000

ID AAT89000 standard; cDNA: 1141 BP.

XX

AC AAT89000;

XX

DT 22-APR-1998 (first entry)

XX Human maspin promoter and partial coding cDNA.

DE Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

KW tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX Homo sapiens.

XX Key

PH Location/Qualifiers

FT promoter

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Db 321 TCAACTGTGACCAAGGCTGCCAGAGCTGGAG 290
 |||| |||| | ||||| ||||| |||

RESULT 8

AAK37539
 ID AAK37539 standard; DNA: 527 BP.

XX
 AC AAK37539;

XX
 DT 06-NOV-2001 (first entry)

XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 12096.

XX
 KW Human: bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX
 OS Homo sapiens.

XX
 PN WO200157276-A2.

XX
 PD 09-AUG-2001.

XX
 PF 30-JAN-2001; 2001WO-US00668.

XX
 PR 04-FEB-2000; 2000US-0180312.

XX
 PR 26-MAY-2000; 2000US-0207456.

XX
 PR 30-JUN-2000; 2000US-0608408.

XX
 PR 03-AUG-2000; 2000US-0632366.

XX
 PR 21-SEP-2000; 2000US-0234687.

XX
 PR 27-SEP-2000; 2000US-0236359.

XX
 PR 04-OCT-2000; 2000GB-0024263.

XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
 WPI: 2001-488900/53.

XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 12096; 658pp + Sequence Listing; English.

XX
 SQ Sequence 527 BP; 101 A; 112 C; 204 G; 110 T; 0 other;

Query Match 50.9%; Score 22.4; DB 22; Length 527;
 Best Local Similarity 81.2%; Pred. No. 14;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 tcaattgtgacaaagctgccagagcttgag 38
 |||| |||| | ||||| ||||| |||

Db 228 tcaactgtgaccaggctgccagagctggag 259

RESULT 9

AAS46482/c
 ID AAS46482 standard; DNA: 12426 BP.

XX
 AC AAS46482;

XX
 DT 18-DEC-2001 (first entry)

XX
 DE Tumour suppressor gene derived chemically modified sequence #204.

XX

KW

Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 cytosine methylation; ds.

XX
 OS Homo sapiens.

XX
 PN WO200168912-A2.

XX
 PD 20-SEP-2001.

XX
 PF 15-MAR-2001; 2001WO-EP02955.

XX
 PR 15-MAR-2000; 2000DE-1013847.

XX
 PR 07-APR-2000; 2000DE-1019058.

XX
 PR 30-JUN-2000; 2000DE-1019173.

XX
 PR 01-SEP-2000; 2000DE-1032529.

XX
 PR 01-SEP-2000; 2000DE-1043826.

XX
 PA (EPIG-) EPIGENOMICS AG.

XX
 PI Olek A, Piepenbrock C, Berlin K;

XX
 WPI: 2001-602752/68.

XX
 PS Claim 1; SEQ ID No 204; 27pp; English.

XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

Query Match 50.9%; Score 22.4; DB 22; Length 12426;
 Best Local Similarity 72.5%; Pred. No. 30;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 tctctttcaattgtggagcagctgccagagcttgagta 40.

Db 7593 TTCTTTCATTATTAACAACACTACCAAAAAAATA 7554
 | ||||| |||| | ||||| |||| | |||| |

RESULT 10

ABL26459/c

ID ABL26459 standard; DNA: 2280 BP.

```
XX AC ABL26459;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 30850.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX PN Drosophila melanogaster.
XX PD WO200171042-A2.
XX PF 27-SEP-2001.
XX PR 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 30850; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 2280 BP; 587 A; 757 C; 451 G; 485 T; 0 other;

Query Match 50.5%; Score 22.2; DB 23; Length 2280;
Best Local Similarity 69.8%; Pred. No. 24;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 1 tctcttcaattgtgacaagctgccaaaggcttgagtagga 43
    |||| |||| | ||| | |||| |||| | ||||
Db 1537 TCTCATTGATCGGGAGAGGATGCCAGATGGGTGGATAGGA 1495

RESULT 11
ABL26458
ID ABL26458 standard; DNA; 4530 BP.
XX AC ABL26458;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 30847.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX PN Drosophila melanogaster.
XX PD WO200171042-A2.
```

```
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 30847; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 4530 BP; 1094 A; 946 C; 1239 G; 1251 T; 0 other;

Query Match 50.5%; Score 22.2; DB 23; Length 4530;
Best Local Similarity 69.8%; Pred. No. 29;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 1 tctcttcaattgtgacaagctgccaaaggcttgagtagga 43
    |||| |||| | ||| | |||| |||| | ||||
Db 1744 tctcattcagatcgaggagagatgccaaagatgggtgtagga 1786

RESULT 12
ABL02384
ID ABL02384 standard; cDNA; 8642 BP.
XX AC ABL02384;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1634.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ss.
XX PN Drosophila melanogaster.
XX PD WO200171042-A2.
XX PF 27-SEP-2001.
XX PR 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
```



```

DR P-PSDB: ABB58281.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
XX Claim 1; SEQ ID NO 1634; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 8642 BP; 2272 A; 1990 C; 1984 G; 2396 T; 0 other;
SQ
Query Match 50.5%; Score 22.2; DB 23; Length 8642;
Best local Similarity 69.8%; Pred. NO. 33;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1 tctcttcgaattgtgacaaagctgccaaagagcgttgagtagga 43
Db 8457 tctcattcagatcgaggagagatgccaaagatgggtgtagga 8499
RESULT 13
AAS16905
ID AAS16905 standard; DNA; 74586 BP.
XX
AC AAS16905;
XX
XX 25-FEB-2002 (first entry)
XX
XX Genomic DNA encoding human transporter polypeptide.
XX
XX Human; transporter; gene therapy; nucleic acid hybridisation; ds;
KW Ionotropic glutamate receptor subfamily; single nucleotide polymorphism;
KW chromosome 9.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1651..69740
FT CDS
FT /*tag= a
FT /product= "Human transporter protein"
FT 1651..2349
FT /*tag= b
FT /number= 1
FT 2350..29157
FT /*tag= c
FT /number= 1
FT /cons_splice= (5'site:NO, 3'site:YES)
FT variation replace(3248,A)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(9928,A)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(11387,C)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(11578,T)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(11731,G)
FT /*tag= h
FT

```

```

/standard_name= "Single nucleotide polymorphism"
/*tag= i
/standard_name= "Single nucleotide polymorphism"
replace(14437,C)
/*tag= j
/standard_name= "Single nucleotide polymorphism"
replace(16732,T)
/*tag= k
/standard_name= "Single nucleotide polymorphism"
replace(18612,C)
/*tag= l
/standard_name= "Single nucleotide polymorphism"
replace(18968,G)
/*tag= m
/standard_name= "Single nucleotide polymorphism"
replace(20360,G)
/*tag= n
/standard_name= "Single nucleotide polymorphism"
replace(23731,A)
/*tag= o
/standard_name= "Single nucleotide polymorphism"
replace(26282,T)
/*tag= p
/standard_name= "Single nucleotide polymorphism"
replace(29047,G)
/*tag= q
/standard_name= "Single nucleotide polymorphism"
29158..29762
/*tag= r
/number= 2
replace(29346,T)
/*tag= s
/standard_name= "Single nucleotide polymorphism"
replace(29542,G)
/*tag= t
/standard_name= "Single nucleotide polymorphism"
replace(29577,G)
/*tag= u
/standard_name= "Single nucleotide polymorphism"
29763..44435
/*tag= v
/number= 2
replace(29779,T)
/*tag= w
/standard_name= "Single nucleotide polymorphism"
replace(32135,T)
/*tag= x
/standard_name= "Single nucleotide polymorphism"
replace(33150,T)
/*tag= y
/standard_name= "Single nucleotide polymorphism"
replace(35710,A)
/*tag= z
/standard_name= "Single nucleotide polymorphism"
replace(37765,A)
/*tag= aa
/standard_name= "Single nucleotide polymorphism"
replace(38468,A)
/*tag= ab
/standard_name= "Single nucleotide polymorphism"
replace(38915,A)
/*tag= ac
/standard_name= "Single nucleotide polymorphism"
replace(39464,C)
/*tag= ad
/standard_name= "Single nucleotide polymorphism"
replace(41195,A)
/*tag= ae
/standard_name= "Single nucleotide polymorphism"
44436..45483
/*tag= af
/number= 3

```

FT	variation	replace(44478,C)	FT	variation	replace(71332,T)
FT		/*tag= ag	FT		/*tag= be
FT	intron	/standard_name= "Single nucleotide polymorphism"	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		45484..52584	FT		replace(72153,G)
FT		/*tag= ah	FT		/*tag= bf
FT		/number= 3	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(51524,G)	FT		replace(72711,K)
FT		/*tag= ai	FT		/*tag= bg
FT	exon	/standard_name= "Single nucleotide polymorphism"	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		52585..52730	FT		replace(74434,C)
FT		/*tag= aj	FT		/*tag= bh
FT		/number= 4	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	intron	52731..57554	XX		
FT		/*tag= ak	PN	WO200181413-A2.	
FT		/number= 4	XX		
FT	variation	replace(54016,T)			
FT		/*tag= al			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(54405,A)			
FT		/*tag= am			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(55007,C)			
FT		/*tag= an			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(55156,G)			
FT		/*tag= ao			
FT	exon	/standard_name= "Single nucleotide polymorphism"			
FT		57555..57670			
FT		/*tag= ap			
FT		/number= 5			
FT	intron	57671..59239			
FT		/*tag= aq			
FT		/number= 5			
FT	exon	59240..59391			
FT		/*tag= ar			
FT		/number= 6			
FT	intron	59392..63553			
FT		/*tag= as			
FT		/number= 6			
FT	exon	63554..63718			
FT		/*tag= at			
FT		/number= 7			
FT	intron	63719..65069			
FT		/*tag= au			
FT		/number= 7			
FT	variation	replace(64177,C)			
FT		/*tag= av			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	exon	65070..65149			
FT		/*tag= aw			
FT		/number= 8			
FT	intron	65150..69400			
FT		/*tag= ax			
FT		/number= 8			
FT		/cons_splice= (5'site:NO, 3'site:NO)			
FT	variation	replace(66196,C)			
FT		/*tag= ay			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(66780,G)			
FT		/*tag= az			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(69176,T)			
FT		/*tag= ba			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	exon	69401..69737			
FT		/*tag= bb			
FT		/number= 9			
FT	variation	replace(70027,R)			
FT		/*tag= bc			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(70419,G)			
FT		/*tag= bd			
FT		/standard_name= "Single nucleotide polymorphism"			

Query Match

50.5%; Score 22.2; DB 24; Length 74586;

Best Local Similarity 69.8%; Pred. No. 54;

Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy

1 tctcttcaattgtgcaagctgccaagagcttgtagcttagtagga 43

11 11111 11111 11 11111 11111 11111 1

Db

4029 tctcttcatctttagattgtttgccaagagcttctaagtagca 4071

RESULT 14

AAS71982

ID AAS71982 standard; cDNA; 1318 BP.

XX

AC AAS71982;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #7786.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR

23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR

P-PSDB; ABG07795.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX

PS Claim 1; SEQ ID No 7786; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

us-09-617-174b-1_copy_914_957.rng

Wed Sep 25 08:14:20 2002

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:15 : Search time 3309.28 Seconds
(without alignments)
179.435 Million cell updates/sec

Title: US-09-617-174B-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagagccttgtagag 44

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estcnu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estl:*
10: gb_estc2:*
11: gb_hnc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	54.5	314	12	BH083050 RPCI-24-2
2	23.6	53.6	385	12	AZ467314
3	23.6	53.6	605	12	A2449029
4	23.6	53.6	676	12	A2462769
5	23.4	53.2	186	12	CNS03C23
6	23.2	53.7	497	9	AA539675
7	23	52.3	515	12	A2144680
8	22.8	51.8	248	9	BB544774
9	22.8	51.8	471	9	AV720858
10	22.8	51.8	477	9	AW399699
11	22.8	51.8	492	12	AZ077326
12	22.8	51.8	507	10	BG137080
13	22.8	51.8	522	12	A0764541
14	22.8	51.8	545	12	AZ748037
15	22.8	51.8	647	12	BH358813
16	22.8	51.8	658	9	BB615382
17	22.8	51.8	686	12	BH110962

C 18	22.8	51.8	750	10	BG16583	BG16583	602677839
C 19	22.8	51.8	971	9	AL523317	AL523317	
C 20	22.8	51.8	2396	11	BC022504	BC022504	Home sep1
C 21	22.6	51.4	380	10	BF196527	BF196527	248303 MA
C 22	22.6	51.4	513	9	A1340928	A1340928	q082A05.x
C 23	22.6	51.4	804	10	BG873257	BG873257	602794102
C 24	22.4	50.9	526	12	B73112	B73112	RPCI11-11E2
C 25	22.4	50.9	831	10	BF965281	BF965281	602268196
C 26	22.4	50.9	840	12	AZ207828	AZ207828	SP_0135-A
C 27	22.4	50.9	856	10	B1262248	B1262248	602954030
C 28	22.4	50.9	1149	12	CNS02K00	CNS02K00	
C 29	22.2	50.5	351	12	A0777266	A0777266	HS_2261_A
C 30	22.2	50.5	405	12	A0634417	A0634417	RPCI-11-4
C 31	22.2	50.5	533	12	A0482843	A0482843	RPCI-11-2
C 32	22.2	50.5	557	10	BC460704	BC460704	RS743326
C 33	22.2	50.5	585	10	BM008659	BM008659	603618528
C 34	22.2	50.5	620	12	A2020356	A2020356	RPCI-23-2
C 35	22.2	50.5	657	10	B1915756	B1915756	603184267
C 36	22.2	50.5	699	10	B1194845	B1194845	602943980
C 37	22.2	50.5	701	10	BF534836	BF534836	602048373
C 38	22.2	50.5	719	10	B1764717	B1764717	603051054
C 39	22.2	50.5	727	10	BE799125	BE799125	601588844
C 40	22.2	50.5	762	12	BH187319	BH187319	033-K_14-
C 41	22.2	50.5	762	12	CNS07RLJ	CNS07RLJ	
C 42	22.2	50.5	801	10	BE736747	BE736747	T7 end of
C 43	22.2	50.5	803	10	B1823154	B1823154	603039596
C 44	22.2	50.5	806	10	B1838085	B1838085	603086668
C 45	22.2	50.5	815	10	BF536852	BF536852	602048440

ALIGNMENTS

RESULT 1	BH083050	314 bp	DNA	linear	GSS 18-JUL-2001
LOCUS	RPCI-24-236D16.TJ				
DEFINITION	BH083050.1				
ACCESSION	BH083050				
VERSION	BH083050.1				
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerogonathi; Muridae; Murinae; Mus.				
AUTHORS	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintel,B., Levins,M., Tsegaye,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.				
TITLE	Mouse BAC End Sequences from Library RPCI-24				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html Plate: 236 Row: D Column: 16 Seq primer: Sp6 Class: BAC ends.				
FEATURES	Location/Qualifiers				
source	1..314				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCI-24-236D16"				
	/clone_lib="RPCI-24"				

```
/sex="Male"
/cell_type="Spleen/Brain"
/Note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RCI-24 Mouse BAC Library produced by Pletier de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      71 a      34 c      62 g      147 t
ORIGIN

Query Match      54.5%; Score 24; DB 12; Length 314;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttcaattgtgacaagctgcagaagagcttgagtagag 44
    ||||| ||||| ||||| ||||| ||||| |||||
Db 60 TTTCATTTTGGAGAGGTATCATGAGCTGTGACAGAG 99

RESULT 2
LOCUS      A2467314      385 bp      DNA      linear      GSS 04-OCT-2000
DEFINITION clone UUGC1M0278A11 R, DNA sequence.
ACCESSION  A2467314
VERSION     A2467314.1 GI:10625439
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 385)
REFERENCE   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS     Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0278 row: A column: 11
            Seq primer: CACACAGAAACAGCATATGACC
            Class: plasmid ends
            High quality sequence stop: 385.
            Location/Qualifiers
                1..385
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0278A11"
                /clone_1lb="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMDA2nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydridynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
```

```
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114/9b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      100 a      94 c      58 g      133 t
ORIGIN

Query Match      53.6%; Score 23.6; DB 12; Length 385;
Best Local Similarity 76.3%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 ttcaattgtgacaagctgcagaagagcttgagtaga 43
    ||||| ||||| ||||| ||||| ||||| |||||
Db 191 TTTCACGTGATCACTGCCAATATGCTTGATAGGA 228

RESULT 3
LOCUS      A2449029      605 bp      DNA      linear      GSS 04-OCT-2000
DEFINITION clone UUGC1M0247N06 F, DNA sequence.
ACCESSION  A2449029
VERSION     A2449029.1 GI:10602411
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 605)
REFERENCE   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
AUTHORS     Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0247 row: N column: 06
            Seq primer: CGTTGTAAACGACGCCACAT
            Class: plasmid ends
            High quality sequence stop: 605.
            Location/Qualifiers
                1..605
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0247N06"
                /clone_1lb="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMDA2nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydridynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
```

```

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g114732114|g14124072.1), a copy number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-gold (Stratagene) cells and selected for ampicillin resistance."
a      140 c      87 g      132 t
acaaagctgcagaagagccttaagtcagag 44
1111111111111111111111111111
AGCAAGTACCATGATGAGCTGTGAGGAAG 568

676 bp      DNA      linear      GSS 04-OCT-2000
JF Mouse 10kb plasmid UGCGM library Mus musculus genomic
CG1M0271E17 F, DNA sequence.

1.1 GI:10620894

use.
ulus.
a. Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
s 1 to 676)
Aoyagi,A., Barber,M., Beacro,T., Duval,B., Hamll,C.,
Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Kelly
e,M., Kose,R., Stokes,R., Tingey,A., von Niederhausern,A.
ht,D., Weiss,R.
ible genome scaffolding with paired end reads from 10kb
Inserts
nshed (2000)
Robert B. Weiss
y of Utah Genome Center
ty of Utah
Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
SA
585 5606
585 7177
dunneGenetics.utah.edu
length: 10000 Std Error: 0.00
2721 row: E column: 17
er: CGTGTAAACGACGCGCACT
lasmid ends
lly sequence stop: 676.
Location/Qualifiers
1. 676
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG1M0271E17"
/clone_1lp="Mouse 10kb plasmid UGCGM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD22nv. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The

```

Query Match	53.28;	Score 23.4;	DB 12;	Length 186;
Best Local Similarity	73.28;	Pred. NO. 69;		

ORGANISM	REFERENCE
<i>Homo sapiens</i>	1 (bases 1 to 471)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

```

/organism="Lycopersicon pennellii"
/db_xref="taxon:28526"
/clone="CLP18B12"
/clone_lib="L. pennellii trichome, Cornell University"
/tissue_type="trichome"

```

```

/dev_stage="mixed stages"
/lab_host="SOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells highly enriched for trichomes, with minor
contamination by other types of leaf cells."
BASE COUNT      152 a      92 c      102 g      131 t
ORIGIN

Query Match      51.8%; Score 22.8; DB 9; Length 477;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 3 tcttcaattgtgacaaagctgccaaagagcttgagtagag 44
      |||||  ||  ||  |||||  |||||  ||
Db 93 TTTTCAATTGGGGAATAAAGCTAGAGCATGAGTAGAG 134

RESULT 11
AZ077326      492 bp      DNA      linear      GSS 31-MAR-2000
LOCUS      RPCI-23-399N3.TJ RPCI-23 Mus musculus genomic clone RPCI-23-399N3,
DEFINITION      DNA sequence.
ACCESSION      AZ077326
VERSION      AZ077326.1 GI:7370225
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 492)
AUTHORS      Zhao, S., Nleman, W., Feldlyum, T., Malek, J., Shamsan, S., Akhret,
      B., Levins, M., Mcgann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
      and Fraser, C.M.
      Mouse BAC End Sequences from Library RPCI-23
      Unpublished (1999)
      Other-GSSs: RPCI-23-399N3.TJ
      Contact: Shaying Zhao
      Department of Eukaryotic Genomics
      The Institute for Genomic Research
      9712 Medical Center Dr., Rockville, MD 20850, USA
      Tel: 301 838 0200
      Fax: 301 838 0208
      Email: szhaob@tigr.org
      Clones are derived from the mouse BAC library RPCI-23. For BAC
      library availability, please contact Pieter de Jong
      (pieter@dejong.med.buflalo.edu). Clones may be purchased from
      BACPAC Resources (http://bacpac.med.buflalo.edu/orderingframe.html)
      or from Resea ch Genetics (info@resgen.com). BAC end page:
      http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
      Plate: 399 row: N column: 3
      Seq primer: SP6
      Class: BAC ends.
FEATURES
      source
      Location/Qualifiers
          1..492
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-399N3"
            /clone_lib="RPCI-23"
            /sex="Female"
            /lab_host="DH10B"
            /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI MethyIase. Size
            selected DNA was cloned into the pBAC3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
```

```

Query Match      51.8%; Score 22.8; DB 12; Length 492;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 2 ccttcaattgtgacaaagctgccaaagagcttgagtaga 43
      ||  ||  |||||  ||  |||||  |||||  ||
Db 238 CTCAGTCCGTGTGACATGTAGCCACAGCTCTGAGCATGA 197

RESULT 12
BG137080      507 bp      mRNA      linear      EST 31-JAN-2001
LOCUS      EST477522 wild tomato pollen Lycopersicon pennellii cDNA clone
DEFINITION      cLPP4P19 5' sequence, mRNA sequence.
ACCESSION      BG137080
VERSION      BG137080.1 GI:12637268
KEYWORDS      EST.
SOURCE      Lycopersicon pennellii.
ORGANISM      Lycopersicon pennellii
REFERENCE      1 (bases 1 to 507)
AUTHORS      van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T.,
      Hansen, C., Rønning, C. and Tanksley, S.
      Generation of ESTs from wild tomato (L. pennellii) pollen
      Unpublished (2001)
      Contact: CGR
      Clemson University Genomics Institute
      100 Jordan Hall, Clemson, SC 29634, USA
      Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES
      source
          1..507
            /organism="Lycopersicon pennellii"
            /cultivar="TA56"
            /db_xref="taxon:28526"
            /clone="cLPP4P19"
            /clone_lib="wild tomato pollen"
            /tissue_type="pollen"
            /dev_stage="SOLR"
            /lab_host="SOLR"
            /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
            XhoI; Pollen was collected from open flowers from
            L.pennellii TA56, and stored at -80 C until library
            construction."
```

```

BASE COUNT      164 a      95 c      122 g      126 t
ORIGIN

Query Match      51.8%; Score 22.8; DB 10; Length 507;
Best Local Similarity 71.4%; Pred. No. 1.6e+02;
Matches 30; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 3 tcttcaattgtgacaaagctgccaaagagcttgagtagag 44
      |||||  ||  ||  |||||  |||||  ||
Db 37 TTTTCAATTGGGGAATAAAGCTAGAGCATGAGTAGAG 78

RESULT 13
AQ764541      522 bp      DNA      linear      GSS 28-JUL-1999
LOCUS      HS_3168.AL.H04_MR CIT Approved Human Genomic Sperm Library D Homo
DEFINITION      sapiens genomic clone Plate=3168 Col=7 Row=O, DNA sequence.
ACCESSION      AQ764541
VERSION      AQ764541.1 GI:5642657
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      1 (bases 1 to 522)
AUTHORS      Human Genome Project
      International Human Genome Mapping
      Consortium
      The Wellcome Trust
      Wellcome Genome Campus
      Hinxton, Cambridgeshire CB3 0ET, UK
      Email: hgp@wghm.ac.uk
      URL: http://www.wghm.ac.uk/
```

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC

```

/organism="Rattus norvegicus"
/strain="BN/SsNsd/MCw"

```

```
Location/Qualifiers
1. .647
/organism="Rattus norvegicus"
/strain="BN/SsNHsd/MCw"
/db_xref="taxon:10116"
```

```
/clone="CH230-163024"  
/clone.lib="CHORI-230 Segment 1"  
/sex="Female"  
/cell_type="Brain"  
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pleter de Jong"  
BASE COUNT      211 a      119 c      132 g      185 t  
ORIGIN
```

```
Query Match      51.8%; Score 22.8; DB 12; Length 647;  
Best Local Similarity 79.4%; Pred. No. 1.7e+02;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 tctcttcaattgtgacaagctgcacaagagct 34  
    ||| ||||| || ||| ||||| ||||| |||  
Db 488 TCTATTTCATTTTGGTCAAGCAGCCAAAGGAT 455
```

Search completed: September 24, 2002, 14:45:19
Job time: 16654 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 13:52:20 ; Search time 1872.72 Seconds
(without alignments)
12749.999 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagcagagaag.....aggataactgtgactccagg ll141

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	805	70.6	170269	9	AC090307
2	805	70.6	173131	2	AC036176
3	805	70.6	173836	2	AP001404
4	337	29.5	214212	2	AC015793
5	226	19.8	214212	2	AC015793
6	89	7.8	1289	9	BC020713
7	61	5.3	2566	6	AX34331
8	61	5.3	2566	9	HSU04313
9	61	5.3	2584	6	AR036974
10	61	5.3	2584	6	AR070462
11	61	5.3	2584	6	115826
12	61	5.3	2584	6	174315
13	60	5.3	164512	9	AL365181
14	57	5.0	179307	2	AC013787
15	57	5.0	208777	2	AC013391
16	55	4.8	158427	9	AC002553
17	53	4.6	2137	6	AX207159
18	52	4.6	140416	2	AC093024
19	52	4.6	156795	2	AC022290
20	52	4.6	172272	2	AC090671
21	52	4.6	182084	2	AC026615
22	50	4.4	81410	9	AP004715
23	50	4.4	148555	2	AP004711
24	50	4.4	160887	2	AC021498
25	50	4.4	207739	2	AC012564
26	49	4.3	12426	6	AX251236
27	48	4.2	31720	9	HSL196E3
28	48	4.2	172800	2	AL356597
29	48	4.2	183430	9	AL391827
30	48	4.2	196476	2	AC055890
31	48	4.2	199255	2	AC084195
32	47	4.1	151791	2	AC093532
33	47	4.1	159906	2	AL590706
34	47	4.1	172004	9	AC008482
35	47	4.1	196606	9	AP000779
36	47	4.1	211967	2	AC013368
37	47	4.1	225157	2	AC027281
38	47	4.1	240579	2	AL360154
39	47	4.1	245202	2	AC092369
40	46	4.0	165560	9	AC068725
41	46	4.0	231972	9	AC068055
42	45	3.9	101611	9	AC018706
43	45	3.9	107017	9	AC012510
44	45	3.9	154306	2	AC034114
45	45	3.9	158905	9	AC005666

ALIGNMENTS

RESULT	1
AC090307	
LOCUS	AC090307 170269 bp DNA linear PRI 30-DEC-2001
DEFINITION	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.
ACCESSION	AC090307
VERSION	AC090307.7 GI:16041402
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 170269)
TITLE	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL	Homo sapiens chromosome 18, clone RP11-851B10
AUTHORS	Unpublished
REFERENCE	2 (bases 1 to 170269)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Bana, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginche, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

COMMENT

On Jul 14, 2000 this sequence version replaced gl:8117315.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gs.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-851B10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Er-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171265 bases at least Q40
Consensus quality: 172100 bases at least Q30
Consensus quality: 172469 bases at least Q20
Insert size: 172836; sum-of-contigs
Quality coverage: 10.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 68896 contig of 68896 bp in length
68997 95899 contig of 26903 bp in length
96000 119268 contig of 23269 bp in length
119369 141739 contig of 22371 bp in length
141840 152655 contig of 10816 bp in length
152756 159466 contig of 6711 bp in length
159567 163465 contig of 3899 bp in length
163566 166433 contig of 2868 bp in length
166534 170112 contig of 3579 bp in length
170213 172095 contig of 1883 bp in length
172196 173836 contig of 1641 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 68896: contig of 68896 bp in length
* 68897 68996: gap of 100 bp
* 68997 95899: contig of 26903 bp in length
* 95900 95999: gap of 100 bp
* 96000 119268: contig of 23269 bp in length
* 119269 119368: gap of 100 bp
* 119369 141739: contig of 22371 bp in length
* 141740 141839: gap of 100 bp
* 141840 152655: contig of 10816 bp in length
* 152656 152755: gap of 100 bp
* 152756 159466: contig of 6711 bp in length
* 159467 159566: gap of 100 bp
* 159567 163465: contig of 3899 bp in length
* 163466 163565: gap of 100 bp
* 163566 166433: contig of 2868 bp in length
* 166434 166533: gap of 100 bp
* 166534 170112: contig of 3579 bp in length
* 170113 170212: gap of 100 bp
* 170213 172095: contig of 1883 bp in length
* 172096 172195: gap of 100 bp
* 172196 173836: contig of 1641 bp in length.
* Location/Qualifiers
1. 173836

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q22"
/clone="RP11-851B10"

misc_feature 1. 68896
/note="assembly_fragment"
misc_feature 68997. .95899
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 96000. .119288
/note="assembly_fragment"
misc_feature 119369. .141739
/note="assembly_fragment"
misc_feature 141840. .152655
/note="assembly_fragment"
misc_feature 152756. .159466
/note="assembly_fragment"
misc_feature 159567. .163465
/note="assembly_fragment"
misc_feature 163566. .166433
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 166534. .170112
/note="assembly_fragment"
misc_feature 170213. .172095
/note="assembly_fragment"
misc_feature 172196. .173836
/note="assembly_fragment"
BASE COUNT 53047 a 34410 c 33807 g 51570 t 1002 others
ORIGIN

Query Match 70.6%; Score 805; DB 2; Length 173836;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1055; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 82 ccaaatgaagagctgttgggaagacagagagacagagctccacgaagagatttca 141
|||||
Db 150643 CCAATGAAGAGCTGTGGAGACAGGAGACAGAGCGCTCCACGAGAGATTTC 150584
|||||
QY 142 gagcagagctgcgtactccttttttttttttttttttttttttttttttttt 201
Db 150583 GAGCAGAGCTGCTACTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 150524
|||||
QY 202 tacagtgttagctcacgctcactgcagcttgcctccacgctcagtgatctctc 261
|||||
Db 150523 TACAGTGTGTAGCTCAGCGCTCAGCTTGACCTGCCAGGCTCAAGTGATCTCTC 150464
|||||
QY 262 gtctcagcttttccaagtaactggggaccacagcagctatcacccagctaggctattgttt 321
Db 150463 GTCTCAGCTTTCCAAAGTAACTGGGACCACAGCGCATGTCATCCACACATAGGCTATTGTTT 150404
|||||
QY 322 tacatttttttagagatgggggtctcaccatgttgcacaggttgggtctcaaaactcctcctc 381
Db 150403 TACATTTTGTGAGAGTGGGGTCTCAGCATGTTGCCAGGTTGGTCTCAAACTCCTGGG 150344
|||||
QY 382 ctcaagcaatccgctcagctcaacctcccaaatgctgggattacagcgctgagccacg 441
Db 150343 CTCAAGCAATCCGCTCAGCTCAACCTCCCAATGCTGGGATTACGGGCTGAGCCACCG 150284
|||||
QY 442 ggcagggctgagtaactcctaatcacagatttttaaaagaacctcttcgcgccaccat 501
Db 150283 CGCCAGGCTGAGTAATCCTAATCACAGATTTTAAAGAAACTTCTCGCCACCACAT 150224
|||||
QY 502 taacaatatctcctaccacaaatttgtagtaataatttgcgtaataatatttttag 561
Db 150223 TAAACAATATCTCTCCACCAATTTGGTAGTAATAATATTGCTAATAGTACCTAATTTTAG 150164
|||||
QY 562 gtaggcactgttttatcatatcatcattcctcttttttgattgtcttctgtttaa 621
Db 150163 GTAGGCACCTGTGTTATACATATATCCATTCCTTTTGTGATTGCTCTTTCTGTTTAAAT 150104
|||||
QY 622 gggcagctacctctcttggtcattctagcaatgagctgctgagtttacacaaaaagaat 681
Db 150103 GGGCAGCTACCTCTCTTGGCATCTAGCAGATGAGCTGCTGAGTTACACAAAAGAAAT 150044
|||||
QY 682 ggagatcagagtaactttttgtgccaccaacgctgctgagaaaatttgcagtttactatata 741
Db 150043 GGAGATCAGAGTACTTTTGTGCCACCAACGCTGCTGAGAAAATTGTAGTGTACTACTTA 149984
|||||


```
* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length
* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp
* 34851 35578: contig of 728 bp in length
* 35579 35678: gap of 100 bp
* 35679 36393: contig of 715 bp in length
* 36394 36493: gap of 100 bp
* 36494 37203: contig of 710 bp in length
* 37204 37303: gap of 100 bp
* 37304 37965: contig of 662 bp in length
* 37966 38065: gap of 100 bp
* 38066 38781: contig of 716 bp in length
* 38782 38881: gap of 100 bp
* 38882 39590: contig of 709 bp in length
* 39591 39690: gap of 100 bp
* 39691 40403: contig of 713 bp in length
* 40404 40503: gap of 100 bp
* 40504 41238: contig of 735 bp in length
* 41239 41338: gap of 100 bp
* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length
```

```
* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 55085 55184: gap of 100 bp
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp
* 57809 58579: contig of 771 bp in length
* 58580 58679: gap of 100 bp

Query Match      29.5%; Score 337; DB 2: Length 214212;
Best Local Similarity 99.3%; Pred. No. 1.2e-181;
Matches 537; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  82 ccaaatgaagaagctgtgggaagacagaggacagaagaacagcgtccacgaagagatttca 141
    |||
Db 168524 CCAAAATGAAGAAGCTGTGGGAAGACAGAGGAGACAAGACAGGCTCCACGAACAGATTTC 168465

QY  142 gagcagagctgcgtactcctcttttttttttttttttttttttttttttttttttttttt 201
    |||
Db 168464 GAGCAGAGCTGCGTACTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTC 168405

QY  202 tacagtgttagctcacgctcactgcagcttttgacctccacagcgtcaagtgcctctc 261
    |||
Db 168404 TACAGTGGTTAGCTCACGCTCAGCTGACGCTTTTGACCTCCCAAGGCTCAAGTGATCCTCTC 168345

QY  262 gtctcagctttccaagtaactggggaccacagcagcatgcacacagcgtaggctattgttt 321
    |||
Db 168344 GTCTCAGCTTTCCAAGTAAC TGGGACACAGCAGCATGCATCACCACACATAGGCTATTGTTT 168285

QY  322 tacatttttttagagatggggctcaccatgttcccaagttgtgtctcaactcctggg 381
    |||
Db 168284 TACATTTTTTTGAGAGATGGGGTCTCACCATGTTGCCAGGTTGGTCTCAAACTCCTGGG 168225

QY  382 ctcaagcaatccgctcacgtcaacctccccaaatgctgggattacagcgctgagccaccg 441
    |||
Db 168224 CTCGAAGCAATCCGCTCAGCTCAACCTCCCAAAATGCTGGGATTACAGGCGTAGCCACCT 168165

QY  442 ggcagggtgagtaactcctaataacagagatttttaaaagaaccttcctgcgccacctat 501
    |||
Db 168164 CGCCAGGCGTGAGTAATCTTAATCAGAGATTTTAAAAAGAAATCTTCTGCGCCACCCAT 168105

QY  502 taacaataatcctcaccacaaatttgtagtaataattttgctaataagtaacctaatattttag 561
    |||
Db 168104 TAAACAATATCTCTTACCATAATTTGGTAGTAAATATATTTTGTCTAATAGTACCTAATTTTATAG 168045

QY  562 gtaggcactgtgttatcatatataccattcctcttttttttttttttttttttttttttttt 621
    |||
Db 168044 GTAGGCACTGTGTTTATACATATATATCCATTCCTTCTTTTGTGATGCTCTTCTGTTTAAAT 167985

QY  622 g 622
Db 167984 G 167984
```

RESULT 5

AC015793

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC015793 Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.

AC015793.2 GI:7144769
HTG; HTGS_PHASE0.
human.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens chromosome 18, clone RP11-1117D15
Unpublished
2 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collange,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D_15

TITLE
JOURNAL
COMMENT

NOTE: This record contains 240 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 695: contig of 695 bp in length
696 795: gap of 100 bp
796 1513: contig of 718 bp in length
1514 1613: gap of 100 bp
1614 2337: contig of 724 bp in length
2338 2437: gap of 100 bp
2438 3161: contig of 724 bp in length
3162 3261: gap of 100 bp
3262 3966: contig of 705 bp in length
3967 4066: gap of 100 bp
4067 4772: contig of 706 bp in length
4773 4872: gap of 100 bp
4873 5564: contig of 692 bp in length
5565 5664: gap of 100 bp
5665 6397: contig of 733 bp in length
6398 6497: gap of 100 bp
6498 7202: contig of 705 bp in length
7203 7302: gap of 100 bp
7303 8028: contig of 726 bp in length
8029 8128: gap of 100 bp
8129 8803: contig of 675 bp in length
8804 8903: gap of 100 bp
8904 9613: contig of 710 bp in length
9614 9713: gap of 100 bp
9714 10409: contig of 696 bp in length
10410 10509: gap of 100 bp
10510 11237: contig of 728 bp in length
11238 11337: gap of 100 bp
11338 12043: contig of 706 bp in length

12044 12143: gap of 100 bp
12144 12867: contig of 724 bp in length
12868 12967: gap of 100 bp
12968 13660: contig of 693 bp in length
13661 13760: gap of 100 bp
13761 14489: contig of 729 bp in length
14490 14589: gap of 100 bp
14590 15314: contig of 725 bp in length
15315 15414: gap of 100 bp
15415 16129: contig of 715 bp in length
16130 16229: gap of 100 bp
16230 16946: contig of 717 bp in length
16947 17046: gap of 100 bp
17047 17769: contig of 723 bp in length
17770 17869: gap of 100 bp
17870 18574: contig of 705 bp in length
18575 18674: gap of 100 bp
18675 19379: contig of 705 bp in length
19380 19479: gap of 100 bp
19480 20186: contig of 707 bp in length
20187 20286: gap of 100 bp
20287 20980: contig of 694 bp in length
20981 21080: gap of 100 bp
21081 21779: contig of 699 bp in length
21780 21879: gap of 100 bp
21880 22594: contig of 715 bp in length
22595 22694: gap of 100 bp
22695 23399: contig of 705 bp in length
23400 23499: gap of 100 bp
23500 24217: contig of 718 bp in length
24218 24317: gap of 100 bp
24318 25026: contig of 709 bp in length
25027 25126: gap of 100 bp
25127 25836: contig of 710 bp in length
25837 25936: gap of 100 bp
25937 26637: contig of 701 bp in length
26638 26737: gap of 100 bp
26738 27446: contig of 709 bp in length
27447 27546: gap of 100 bp
27547 28274: contig of 728 bp in length
28275 28374: gap of 100 bp
28375 29095: contig of 721 bp in length
29096 29195: gap of 100 bp
29196 29917: contig of 722 bp in length
29918 30017: gap of 100 bp
30018 30724: contig of 707 bp in length
30725 30824: gap of 100 bp
30825 31529: contig of 705 bp in length
31530 31629: gap of 100 bp
31630 32324: contig of 695 bp in length
32325 32424: gap of 100 bp
32425 33116: contig of 692 bp in length
33117 33216: gap of 100 bp
33217 33936: contig of 720 bp in length
33937 34036: gap of 100 bp
34037 34750: contig of 714 bp in length
34751 34850: gap of 100 bp
34851 35578: contig of 728 bp in length
35579 35678: gap of 100 bp
35679 36393: contig of 715 bp in length
36394 36493: gap of 100 bp
36494 37203: contig of 710 bp in length
37204 37303: gap of 100 bp
37304 37965: contig of 662 bp in length
37966 38065: gap of 100 bp
38066 38781: contig of 716 bp in length
38782 38881: gap of 100 bp
38882 39590: contig of 709 bp in length
39591 39690: gap of 100 bp
39691 40403: contig of 713 bp in length
40404 40503: gap of 100 bp
40504 41238: contig of 735 bp in length
41239 41338: gap of 100 bp

```

* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length
* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 55085 55184: gap of 100 bp
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp
* 57809 58579: contig of 771 bp in length
* 58580 58679: gap of 100 bp

Query Match 19.8%; Score 226; DB 2; Length 214212;
Best Local Similarity 100.0%; Pred. No. 6.8e-118;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 tctttcaattgtgacaagctccaagagcgttgtagtaggagagtgagtgcccgagcg 975
DB 57085 TCTTTCAATTGTGGACAAGCTCCCAAGAGCTTGAGTAGGAGGAGTGCCCGCAGGCG 57144

QY 976 gggcgggcgggcggtgagctggcgtgagtgagcggtgagcggtgagcggtgagcg 1035
DB 57145 GGGCGGGCGGGCGGTGAGCTGGCGTGGCAGTGCGCGGTGCTGCCAGGTGAGC 57204

QY 1036 caccgtgcttctgccacagcagcgtcgctccacatccagctcttctgctcctcgctt 1095
DB 57205 CACCGCTGCTTCTGCCACAGCAGCGTGGCTCCATCCAGGTCTTGTGCTCGCTT 57264

QY 1096 gctgttctctttccacgcattttccaggataaactgtgactccag 1141
DB 57265 GCCTGTCTCTTTTCCACGCAATTTCCAGGATAACTGTGACTCCAGG 57310

RESULT 6
BC020713
LOCUS Homo sapiens, Similar to serine (or cysteine) proteinase inhibitor,
DEFINITION clade B (ovalbumin), member 5, clone MGC:22489 IMAGE:4248249, mRNA,
complete cds.
ACCESSION BC020713
VERSION BC020713.1 GI:18089113
KEYWORDS MGC.

```

```

SOURCE
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 1289)
JOURNAL Strausberg,R.
Direct Submission
Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 35 Row: C Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4505788.
FEATURES
source
Location/Qualifiers
1..1289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:22489 IMAGE:4248249"
/tissue_type="Prostate"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
96..791
/codon_start=1
/product="Similar to serine (or cysteine) proteinase
inhibitor, clade B (ovalbumin), member 5"
/db_xref="GI:18089114"
/translation="MDALQLANSFAVDLFKOLCEKPLGNVLFSPICLSTLSLAOV
GAKGDTANEIGVLFHFENVKDVDFGOTVTSQVAKLSFSYSLKLIKRLYVDSLNLSL
EFISSTRKPYAKELETVDKLETKGQINNSIKDLTDGHFENILADNSVNDQTKIL
VYNAAYFVGKWKMKFPPESETEKCPFRVNVKVCAGCSSRSPDIIDVKNDRDRVGHKSIP
MNLRLRAPAKLS"
BASE COUNT 401 a 274 c 280 g 334 t
ORIGIN

Query Match 7.8%; Score 89; DB 9; Length 1289;
Best Local Similarity 100.0%; Pred. No. 2e-39;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1053 agacacgctgcctccacatccagctcttctgctcctcagctgctgcttcttccac 1112
DB 1 AGACACGGTCGCTCCACATCCAGGTCTTTGTGCTCCTCGCTGCTGCTTCTCTTCCAC 60

QY 1113 gcatcttccaggataactgtgactccag 1141
DB 61 GCATTTTCCAGGATAACTGTGACTCCAGG 89

RESULT 7
AX334331
LOCUS AX334331
DEFINITION Sequence 2566 bp DNA
ACCESSION AX334331
VERSION AX334331.1
KEYWORDS PAT 09-JAN-2002

```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
TITLE 1 (sites)
JOURNAL Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.,
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 4840 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES Location/Qualifiers
source
1..2566
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 786 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2566;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgttcctttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTGCTGCTGCTGCTGCTTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

Qy 1141 g 1141
Db 69 G 69

RESULT 8
HSU04313
LOCUS HSU04313 2566 bp mRNA linear PRI 11-JUN-1994
DEFINITION Human maspin mRNA, complete cds.
ACCESSION U04313
VERSION U04313.1 GI:453368
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
Zou, Z., Anisowicz, A., Neveu, M., Rafidi, K., Sheng, S., Sager, R.,
Hendrix, M.J., Seftor, E. and Thor, A.
TITLE Maspin, a serpin with tumor suppressing activity in human mammary
epithelial cells
JOURNAL Science 263, 526-529 (1994)
MEDLINE 94120413
REFERENCE 2 (bases 1 to 2566)
AUTHORS Anisowicz, A.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1993) Anthony Anisowicz, Cancer Genetics, Dana
Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES Location/Qualifiers
source
1..2566
/organism="Homo sapiens"
/isolate="76N"
/db_xref="taxon:9606"
/clone="232-1"
/sex="female"
/cell_line="76N"
/cell_type="epithelial"
/tissue_type="mammary"
/clone_lib="76N cDNA library in lambda Zap II"
1..75
76..1203
/codon_start=1
/product="maspin"
/protein_id="AAAI8957.1"
/db_xref="GI:453369"
/translation="MDALQLANSAFAVDLFKOLCEKPELGNVLFSPICLSTLSLAQV

GAAGDTANEIGQVLFHFNKVIDPFGQTVTSDVNVKLSSEFYSKLIRLRYDKSLNST
EFISSTKRPYAKELTVDFKDKLEETKQINNSIKDLTDGHEFNILADNSVNDQTKIL
VNAAYVPGWMMKFPSETKECPFRNLKTDTPQVMNMMEATFCMGNIDISNCKIIE
LPQNKHLSMFIILLPKDVEDSTGLKIEKQLNSELISQWTPNSTMANAKVLSIKPKF
KVEKMDIPKACLENLGLKHIFSEDSTDFSGMSETKGVALSNIHKVLCLETTEDGGDSI
EVPGARILQHKDELNADHPFIYIIRHNKTRNIIFFGKFCSP"
misc_feature
1093..1098
/note="putative serpin reactive center"
3'UTR 1201..2566
polyA_signal 2545..2550
BASE COUNT 786 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 9; Length 2566;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgttcctttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTGCTGCTGCTGCTTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

Qy 1141 g 1141
Db 69 G 69

RESULT 9
AR036974
LOCUS AR036974 2584 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5801001.
ACCESSION AR036974
VERSION AR036974.1 GI:5954830
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Zou, Z. and Anisowicz, A.
TITLE Method of detecting cancer
JOURNAL Patent: US 5801001-A 1 01-SEP-1998;
FEATURES Location/Qualifiers
source
1..2584
/organism="unknown"
BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgttcctttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTGCTGCTGCTTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

Qy 1141 g 1141
Db 69 G 69

RESULT 10
AR070462
LOCUS AR070462 2584 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 1 from patent US 5905023.
ACCESSION AR070462
VERSION AR070462.1 GI:7221350
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Zou, Z. and Anisowicz, A.

TITLE Maspin, a serpin with tumor suppressing activity
JOURNAL Patent: US 5905023-A 1 18-MAY-1999;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgctgctgttctctttccacgcacattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTGCCTGTTCCCTTTCCACGCAATTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 11
LOCUS I15826 2584 bp DNA linear PAT 02-APR-1996
DEFINITION Sequence 1 from patent US 5470970.
ACCESSION I15826
VERSION I15826.1 GI:1250734
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager R., Anisowicz A. and Zou Z.
TITLE Maspin, a serpin with tumor suppressing activity
JOURNAL Patent: US 5470970-A 1 28-NOV-1995;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgctgctgttctctttccacgcacattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTGCCTGTTCCCTTTCCACGCAATTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 12
LOCUS I74315 2584 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 13 from patent US 5688641.
ACCESSION I74315
VERSION I74315.1 GI:3010456
KEYWORDS Unknown.
SOURCE Unknown.

ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager R., Zou Z., Lee S., Whan. and Tomasetto C. Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 13 18-NOV-1997;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgctgctgttctctttccacgcacattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTGCCTGTTCCCTTTCCACGCAATTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 13
LOCUS AL365181/c 164512 bp DNA linear PRI 22-JAN-2002
DEFINITION Human DNA sequence from clone RP11-284F21 on chromosome 1, complete sequence.

ACCESSION AL365181
VERSION AL365181.24 GI:18307339
KEYWORDS HTG.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Chapman, J.
Direct Submission

REFERENCE 1 (bases 1 to 164512)
AUTHORS
TITLE
JOURNAL

COMMENT Submitted (22-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 23, 2002 this sequence version replaced gi:18250487.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-284F21 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-284F21. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-284F21 is at 1 in this sequence.
The true left end of clone RP11-66D17 is at 162513 in this sequence. The true right end of clone RP11-98G7 is at 115430 in this sequence.

FEATURES Location/Qualifiers
source 1..164512
/organism="Homo sapiens"
/db_xref="taxon:9606"

	REFERENCE	1 (bases 1 to 179307)	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
	AUTHORS	Homo sapiens chromosome 15, clone RP11-531E13	
	TITLE	Unpublished	
	REFERENCE	2 (bases 1 to 179307)	
	AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boquistavsky,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehocaky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.	
	TITLE	Direct Submission	
	JOURNAL	Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
	COMMENT	On Dec 28, 2001 this sequence version replaced gi:14190734. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html	
		Center: Whitehead Institute/ MIT Center for Genome Research	
		Center code: WIBR	
		Web site: http://www-seq.wi.mit.edu	
		Contact: sequence_submissions@genome.wi.mit.edu	
		----- Project Information	
		Center project name: L3840	
		Center clone name: 531_E_13	

		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 9 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	
		* runs of N, but the exact sizes of the gaps are unknown.	
		* This record will be updated with the finished sequence	
		* as soon as it is available and the accession number will	
		* be preserved.	
		* 1 68483: contig of 68483 bp in length	
		* 68484 68583: gap of 100 bp	
		* 68584 75863: contig of 7280 bp in length	
		* 75864 75963: gap of 100 bp	
		* 75964 106600: contig of 30637 bp in length	
		* 106601 106700: gap of 100 bp	
		* 106701 120099: contig of 13399 bp in length	
		* 120100 120199: gap of 100 bp	
		* 120200 131661: contig of 11462 bp in length	
		* 131662 131761: gap of 100 bp	
		* 131762 134040: contig of 2279 bp in length	
		* 134041 134140: gap of 100 bp	
		* 134141 157957: contig of 23817 bp in length	
		* 157958 158057: gap of 100 bp	
		* 158058 177336: contig of 19279 bp in length	
		* 177337 177436: gap of 100 bp	
		* 177437 179307: contig of 1871 bp in length.	
	FEATURES	Location/Qualifiers	
	SOURCE	1. 179307	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="15"	
		/map="15"	
		/clone="RP11-531E13"	
		/clone_lib="RPC1-11 Human Male BAC"	
		48710 a 40689 c 40017 g 49034 t 857 others	
	BASE COUNT		
	ORIGIN		
		Query Match 5.0%; Score 57; DB 2; Length 179307;	

Best Local Similarity 100.0%; Pred. No. 7.4e-21;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 325 atttttttagagatggggtctcacatgttcccgaggttggtctcctcaactcctggg 381
|||||
Db 57986 ATTTTGTAGAGATGGGGTCTCACCATGTTGCCAGGTGGTCTCAAACTCCTGGG 57930

RESULT 15
AC013391
LOCUS Homo sapiens chromosome 15 clone RP11-429B14 map 15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 8 unordered pieces.
AC013391
AC013391.6 GI:16303477
VERSION HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208777)
Biren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 15, clone RP11-429B14
Unpublished
2 (bases 1 to 208777)
Biren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 20, 2001 this sequence version replaced gi:12643085.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3803
Center clone name: 429_B14

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 281: contig of 281 bp in length
* 282 381: gap of 100 bp
* 382 12491: contig of 12110 bp in length
* 12492 12591: gap of 100 bp
* 12592 14903: contig of 2312 bp in length
* 14904 15003: gap of 100 bp
* 15004 20280: contig of 5277 bp in length
* 20281 20380: gap of 100 bp
* 20381 27441: contig of 7061 bp in length
* 27442 27541: gap of 100 bp

* 27542 42474: contig of 14933 bp in length
* 42475 42574: gap of 100 bp
* 42575 119846: contig of 77272 bp in length
* 119847 119946: gap of 100 bp
* 119947 208777: contig of 88831 bp in length.
location/Qualifiers
1. 208777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-429B14"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 54332 a 50627 c 49063 g 53962 t 793 others
ORIGIN

Query Match 5.0%; Score 57; DB 2; Length 208777;
Best Local Similarity 100.0%; Pred. No. 7.5e-21;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 325 atttttttagagatggggtctcacatgttgcacagttgtccagttgtctcaaacctcctggg 381
|||||
Db 8507 ATTTTGTAGAGATGGGGTCTCACCATGTTGCCAGGTGGTCTCAAACTCCTGGG 8563

Search completed: September 24, 2002, 16:50:27
Job time: 10687 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:25 ; Search time 197.2 Seconds
(without alignments)
9934.074 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagacagcagagaag.....agataactgtactccagg 1141

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	91.1	1141	18 AAT89000	Human maspin promo
2	89	7.8	410	21 AAC00200	Human secreted pro
3	61	5.3	2584	15 AAQ56790	Maspin DNA. Homo
4	61	5.3	2584	17 AAT07120	Tumour suppressor
5	61	5.3	2584	19 AAV55017	Maspin coding sequ
6	61	5.3	2584	20 AAX36183	cDNA encoding huma
7	53	4.6	2136	22 AAH47053	Human maspin cDNA
8	49	4.3	12426	22 AAS46482	Tumour suppressor
9	47	4.1	51	18 AAT89002	Human maspin oligo

10	43	3.8	14175	22 AAS27814	DNA encoding novel
11	43	3.8	14175	22 AAK78859	Human immune/haema
12	43	3.8	23164	22 AAK79678	Human immune/haema
13	43	3.8	26995	22 AAS29963	Human lung antigen
14	40	3.5	4086	21 AAK16677	Human secreted pro
15	40	3.5	12834	22 AAK73037	Human immune/haema
16	39	3.4	32191	22 AAS30497	DNA encoding novel
17	39	3.4	32191	22 AAL06277	Human reproductive
18	39	3.4	46107	22 AAK71730	Human immune/haema
19	38	3.3	8418	22 AAS41906	Genomic sequence #
20	38	3.3	10458	22 AAK77961	Human immune/haema
21	38	3.3	17874	22 AAK71070	Human immune/haema
22	38	3.3	33923	22 AAK67071	Human immune/haema
23	38	3.3	49999	20 AAZ23904	Human LOBO homolog
24	37	3.2	1388	22 AAK63064	Human immune/haema
25	37	3.2	1389	22 AAK69065	Human immune/haema
26	37	3.2	4505	22 AAK64819	Human immune/haema
27	37	3.2	9234	22 AAK89255	Human digestive sy
28	37	3.2	14150	22 AAK64820	Human immune/haema
29	37	3.2	465217	24 ABA90193	Human oestrogen re
30	36	3.2	534	22 AAH09878	Human cDNA clone (
31	36	3.2	1343	22 AAK67011	Human immune/haema
32	36	3.2	1705	22 AAH17516	Human cDNA sequenc
33	36	3.2	1717	22 AAH13941	Human cDNA sequenc
34	35	3.1	288	22 AA186011	Human polynucleoti
35	35	3.1	330	22 AA185941	Human polynucleoti
36	35	3.1	3303	16 AAQ76024	TGF-beta-3 promote
37	35	3.1	3303	22 AAH26121	Human transforming
38	35	3.1	22700	22 AAK70746	Human immune/haema
39	34	3.0	273	22 AAK64746	Human immune/haema
40	34	3.0	273	22 AAK64748	Human immune/haema
41	34	3.0	273	22 AAK64752	Human immune/haema
42	34	3.0	274	22 AAK64741	Human immune/haema
43	34	3.0	274	22 AAK64744	Human immune/haema
44	34	3.0	379	22 AAF67015	Novel human polynu
45	34	3.0	745	22 AAL24371	Human breast cance

ALIGNMENTS

RESULT 1

AAT89000
ID AAT89000 standard; cDNA; 1141 BP.

XX AC AAT89000;

XX DT 22-APR-1998 (first entry)

XX DE Human maspin promoter and partial coding cDNA.

XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

XX KW tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX OS Homo sapiens.

XX FH Key

XX FT promoter

XX FT Location/Qualifiers

XX FT 1..1141

XX FT /tag= a

XX FT /note= "maspin promoter region"

XX FT misc_signal

XX FT 443..449

XX FT /tag= b

XX FT misc_signal

XX FT 451..457

XX FT /note= "Putative regulatory element AP2"

XX FT /tag= c

XX FT misc_signal

XX FT 485..490

XX FT /note= "Putative regulatory element API"

XX FT /tag= d

XX FT misc_signal

XX FT 684..691

XX FT /note= "Putative regulatory element Ets"

XX FT /tag= e

XX FT misc_signal

XX FT 846..851

XX FT /note= "Putative regulatory element HRE"

FT /*tag= f
FT /Note- "Putative regulatory element Ets"
FT 907..913
FT /*tag= g
FT /Note- "Putative regulatory element APl"
FT prim_transcript 958..1141
FT /*tag= h
FT /Note- "partial"
XX
PN WO9736179-A1.
XX
XX 02-OCT-1997.
XX
XX 28-MAR-1997; 97WO-US05186.
XX
XX 28-MAR-1996; 96US-0014368.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX (DAND) DANA FARBER CANCER INST.
XX (PARD/) PARDEE A.
XX
XX Sagar R, Zhang M;
XX WPI; 1997-489785/45.
XX
XX Maspin gene promoter fragment - used to identify compounds for
XX treatment of prostate or breast cancer
XX
XX Claim 4; Fig 3; 51pp; English.

This sequence encodes the human maspin promoter region, including the
transcription start site. Maspin is a serpin which is expressed in
mammary epithelial cells. Its expression in these cells decreases
with increasing malignancy and is lost in during metastasis. Maspin
protein is also known to inhibit the mobility of tumour cells. This gene
can be used in method for screening compounds to identify candidate
compounds for the treatment of prostate cancer, or breast cancer. It can
also be used to identify compounds that increase the expression of
maspin, and for detecting the presence of metastatic prostate epithelial
cells.

XX Sequence 1141 BP; 261 A; 292 C; 277 G; 311 T; 0 other;

Query Match 91.1%; Score 1039; DB 18; Length 1141;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agataagcacagacagagaagcaaccagctcgtttcaggtctcttctcgtgattctg 60
DB 1 agataagcacagacagagaagcaaccagctcgtttcaggtctcttctcgtgattctg 60
QY 61 gctgaaggagtaggtgtccaccacaaatgaagaagctgtggaagacagagagacaagaac 120
DB 61 gctggaaggagtaggtgtccaccacaaatgaagaagctgtggaagacagagagacaagaac 120
QY 121 aggctccacgaagagatttcagagcagagctgcgtactccttcttcttcttcttcttctt 180
DB 121 aggctccacgaagagatttcagagcagagctgcgtactccttcttcttcttcttcttctt 180
QY 181 gctctgtaccacaggtgaagtacagtgttagctcagcaggtcactgcagctttgacctc 240
DB 181 gctctgtaccacaggtgaagtacagtgttagctcagcaggtcactgcagctttgacctc 240
QY 241 ccaggctcaagtgtactctctcgtctcagctttcccaagtaactgggaccacagagctgcat 300
DB 241 ccaggctcaagtgtactctcgtctcagctttcccaagtaactgggaccacagagctgcat 300
QY 301 caccacgttaggctattgttttacatttttttagagatggggtctcaccattgttgccca 360
DB 301 caccacgttaggctattgttttacatttttttagagatggggtctcaccattgttgccca 360
QY 361 ggttggtctcaaaactcctgggctcagaagcaatccgctcagctcaactcccaaatgctgg 420

DB 361 ggttggtctcaaaactcctgggctcaagcaatccgctcagctcaactcccaatgctgg 420
QY 421 gattacaggtcgtgagccacggcgccaggtcgtgagtaatacctaatcacagatttataaa 480
DB 421 gattacaggtcgtgagccacggcgccaggtcgtgagtaatacctaatcacagatttataaa 480
QY 481 gaaactctcgtgcacccattaaacaatatctcctaccatttggtagtaataattttg 540
DB 481 gaaactctcgtgcacccattaaacaatatctcctaccatttggtagtaataattttg 540
QY 541 ctaattagtaacctaatttttaggtaggcactgtttatatacatatcattcctctttt 600
DB 541 ctaattagtaacctaatttttaggtaggcactgtttatatacatatcattcctctttt 600
QY 601 ttgatttctcttctgtttaatgggcagctacctctcttggcatcttagcagaatgagctgc 660
DB 601 ttgatttctcttctgtttaatgggcagctacctctcttggcatcttagcagaatgagctgc 660
QY 661 tgcagtttacacaaaagaatggagatcagagtagcttttttggccacacagtgctgag 720
DB 661 tgcagtttacacaaaagaatggagatcagagtagcttttttggccacacagtgctgag 720
QY 721 aaattttagtgttactatcatcacattacttttatttctcatgaattttcaccttcc 780
DB 721 aaattttagtgttactatcatcacattacttttatttctcatgaattttcaccttcc 780
QY 781 ggtcctcgtggtggccgagagagattgcccagcatgtctgtacgtatgcatgtaactcac 840
DB 781 ggtcctcgtggtggccgagagagattgcccagcatgtctgtacgtatgcatgtaactcac 840
QY 841 agcccttctcgtccgcaacatgttgaggagccttttggagctgtgcagacacagacaact 900
DB 841 agcccttctcgtccgcaacatgttgaggagccttttggagctgtgcagacacacagacaact 900
QY 901 tcagctgaatcatctcttcaattgttgacaagctgccaagagcttgagtaggagagg 960
DB 901 tcagctgaatcatctcttcaattgttgacaagctgccaagagcttgagtaggagagg 960
QY 961 agtgcgcgcagcggcggtggggcggtggagctgggtgggtgggtgggtgggtgggtgggt 1020
DB 961 agtgcgcgcagcggcggtggggcggtggagctgggtgggtgggtgggtgggtgggtgggt 1020
QY 1021 gctgccacaggtgagccacgcgtcttctgccagacacacgctgcgtccacatccaggtct 1080
DB 1021 gctgccacaggtgagccacgcgtcttctgccagacacacgctgcgtccacatccaggtct 1080
QY 1081 ttgtgtcctcgtctgctgttcttcttccacgcatttccagagataactgtgactccag 1140
DB 1081 ttgtgtcctcgtctgctgttcttcttccacgcatttccagagataactgtgactccag 1140
QY 1141 g 1141
DB 1141 g 1141
RESULT 2
AAC00200
ID AAC00200 standard; cDNA; 410 BP.
XX
AC AAC00200;
XX
XX 06-OCT-2000 (first entry)
DE Human secreted protein 5' EST, SEQ ID NO: 198.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EPI033401-A2.
XX

```

PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
DR P-PSDB; AAG00194.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 198; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 410 BP; 103 A; 108 C; 76 G; 122 T; 1 other;
SQ
Query Match 7.8%; Score 89; DB 21; Length 410;
Best Local Similarity 100.0%; Pred. No. 2.9e-34;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1053 agacacggtgcctccacatccaggtcttggctcctcgtcgtcgtcttcctccac 1112
DB 1 agacacggtgcctccacatccaggtcttggctcctcgtcgtcgtcttcctccac 60
QY 1113 gcattttccagataactgtgactccag 1141
DB 61 gcattttccagataactgtgactccag 89
RESULT 3
AAQ56790
ID AAQ56790 standard; DNA; 2584 BP.
XX
XX AAQ56790;
XX
XX 30-SEP-1994 (first entry)
DE Maspin DNA.
XX
XX Maspin; serine protease inhibitor; mammary; epithelial cells; cancer;
KW breast tissue; well-differentiated; poorly differentiated; antibody;
KW primary carcinoma; lymph node lesion; metastatic lesion; carcinoma;
KW tumour; tumour suppression; invasion suppression; anti-maspin;
KW anticancer compound; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 76..1203
FT CDS /*tag= a
FT /product= Maspin
XX

```

```

PN W09405804 -A.
XX
XX 17-MAR-1994.
XX
XX 01-SEP-1993; 93WO-US08322.
XX
XX 01-SEP-1992; 92US-0938823.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Sager R;
PI
XX
XX WPI; 1994-101208/12.
DR P-PSDB; AAR50938.
XX
XX DNA encoding maspin which has tumour suppressing activity - used
PT to produce prods. for the diagnosis, staging and treatment of
PT carcinoma and for screening anti-cancer agents
XX
XX Claim 1; Page 35-37; 67pp; English.
XX
XX This sequence represents the maspin gene which was isolated by
CC subtractive hybridisation. This gene encodes a serine protease
CC inhibitor and is expressed in normal mammary epithelial cells in
CC culture and in normal breast tissue. Expression of this gene
CC decreases during progression from well-differentiated to poorly
CC differentiated primary carcinomas, and is absent in most lymph node
CC and distant metastatic lesions. The maspin protein has tumour
CC suppressing and invasion suppressing activity. The DNA and anti-
CC maspin antibodies can be used to determine whether test cells are
CC cancerous or for determining whether a test carcinoma cell represents
CC an early-stage or advanced metastatic carcinoma. The level of
CC expression of maspin in cells can be used for screening anticancer
CC compounds. Compounds which increase the amount of maspin in, or in
CC the immediate vicinity of carcinoma cells can be used to treat
CC carcinomas. Maspin, and DNA encoding maspin, can also be used in
CC the treatment of carcinomas
XX
XX Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;
SQ
Query Match 5.3%; Score 61; DB 15; Length 2584;
Best Local Similarity 100.0%; Pred. No. 3.1e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1081 ttgtgctcctcgtcgttcctttccacgcattttccagataactgtgactccag 1140
DB 9 ttgtgctcctcgtcgttcctttccacgcattttccagataactgtgactccag 68
QY 1141 g 1141
DB 69 g 69
RESULT 4
AAQ07120
ID AAQ07120 standard; cDNA; 2584 BP.
XX
XX AAQ07120;
XX
XX 15-FEB-1996 (first entry)
DE Tumour suppressor maspin cDNA.
XX
XX Maspin; serpin; tumour suppressor; breast cancer; carcinoma;
KW gene therapy; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 76..1203
FT CDS /*tag= a
XX

```


XX US905023-A.
 XX 18-MAY-1999.
 XX 07-JUN-1995; 95US-0477112.
 XX 01-SEP-1993; 93US-0121714.
 XX 28-FEB-1991; 91US-0662216.
 XX 28-FEB-1992; 92US-0844296.
 XX 01-SEP-1992; 92US-0938823.
 XX 07-JUN-1995; 95US-0477112.
 XX (DAND) DANA FARBER CANCER INST INC.
 XX Anisowicz A, Sager R, Zou Z;
 DR WPI; 1999-326381/27.
 DR P-PSDB; AAY02607.
 XX Inhibiting the growth of a carcinoma using a serpin with tumour
 PT suppressing activity
 XX Claim 1; Columns 17-22; 29pp; English.
 XX The present sequence encodes a human serpin called maspin. The
 CC specification describes a method for inhibiting the growth of
 CC a carcinoma by treating carcinoma cells with maspin. The method
 CC is useful as genetic therapy to help control carcinomas
 CC characterized by down-regulated maspin expression.
 XX Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;
 SQ

Query Match 5.3%; Score 61; DB 20; Length 2584;
 Best Local Similarity 100.0%; Pred. No. 3.1e-20;
 Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1081 ttgtctctcgtctgctgtctctttccacgcattttccaggataactgtgactccag 1140
 Db 9 ttgtctctcgtctgctgtctctttccacgcattttccaggataactgtgactccag 68

Oy 1141 g 1141
 Db 69 g 69

RESULT 7
 AAH47053
 ID AAH47053 standard; DNA; 2136 BP.
 XX AAH47053;
 XX 29-OCT-2001 (first entry)
 DE Human maspin cDNA sequence.
 XX Drug resistance; resistance gene; semaphorin D; B94; mel-14 antigen;
 KW 24p3; proliferin; maspin; cancer; cytostatic; gene therapy; ss.
 XX Homo sapiens.
 OS WO200155455-A2.
 XX 02-AUG-2001.
 XX 31-JAN-2001; 2001WO-US03161.
 XX 31-JAN-2000; 2000US-0179191.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA (JINS/) JIN S.
 XX

PI Jin S;
 XX WPI; 2001-488799/53.
 XX Determining if a compound modulates the drug resistance of a cell,
 PT comprises determining the expression or activity level of a resistance
 PT sequence in a cell in the presence of the test compound -
 XX Example 1; Fig 5A-B; 79pp; English.
 XX The invention relates to a method of determining whether a test compound
 CC modulates the drug resistance of a cell that comprises determining the
 CC expression or activity level of resistance genes (e.g. semaphorin D, B94,
 CC mel-14 antigen, 24p3, proliferin or maspin) in a cell in the presence of
 CC the test compound, and comparing its expression or activity level in a
 CC cell without the test compound. The drug resistant sequences are useful
 CC in identifying drug resistant cells, in screening methods directed to the
 CC identification of compounds that can modulate the drug resistance of a
 CC cell type or multiple cell types. An isolated resistance protein can be
 CC used as an immunogen to generate antibodies that bind the resistance
 CC protein. Resistance nucleic acids may be inserted into vectors and used
 CC as gene therapy vectors. An anti-resistance protein antibody may be used
 CC to isolate a resistance protein, or facilitate the purification of
 CC natural resistance protein from cells and of recombinantly produced
 CC resistance protein expressed in host cells. The methods are useful for
 CC treating a subject having a disorder, such as a drug-resistance cancer,
 CC characterized by aberrant resistance sequence expression or activity by
 CC administering to the subject a resistance modulator. The present sequence
 CC represents a human maspin cDNA sequence, whose expression was decreased
 CC in drug resistant EMT6 tumours.
 XX Sequence 2136 BP; 658 A; 454 C; 428 G; 596 T; 0 other;
 SQ

Query Match 4.6%; Score 53; DB 22; Length 2136;
 Best Local Similarity 100.0%; Pred. No. 3.2e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1089 ctgcgttcctgctgtctctttccacgcattttccaggataactgtgactccag 1141
 Db 87 ctgcgttcctgctgtctctttccacgcattttccaggataactgtgactccag 139

RESULT 8
 AAS46482/C
 ID AAS46482 standard; DNA; 12426 BP.
 XX AAS46482;
 XX 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #204.
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX Homo sapiens.
 OS WO200168912-A2.
 XX 20-SEP-2001.
 XX 15-MAR-2001; 2001WO-EP02955.
 XX 15-MAR-2000; 2000DE-1013847.
 XX 06-APR-2000; 2000DE-1019058.
 XX 07-APR-2000; 2000DE-1019173.
 XX 30-JUN-2000; 2000DE-1032529.
 XX 01-SEP-2000; 2000DE-1043826.
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX

PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
XX Claim 1; SEQ ID NO 204; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (Cp DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with CpG dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes. Sequences with even numbered Seq ID numbers are the
CC complementary sequence of the corresponding odd numbered sequence (e.g.
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

```

Query Match      4.38; Score 49; DB 22; Length'12426;
Best Local Similarity 100.08; Pred. NO. 3.3e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy    733 ttactatcatcacacattacttttatttcacgcgaatatcaccctcg 781
      |||||
Db     7774 TTACTATCATCACACATTACTTTTATTTCATCGAATATTTCACCTCCG 7726

RESULT          9
AAT89002/c
ID   AAT89002 standard; DNA; 51 BP.
XX
XX AC
XX AAT89002;
XX
XX DT 22-APR-1998 (first entry)
XX
XX DE Human maspin oligonucleotide primer OLI DNA.
XX
XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;
KW tumour cell; treatment; prostate cancer; breast cancer; metastasis;
KW primer; ss.
XX
XX OS Synthetic.
OS Homo sapiens.
XX
XX PN WO9736179-A1.
XX
XX PD 02-OCT-1997.
XX
XX PF 28-MAR-1997; 97WO-US05186.
XX

```


PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 1474; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system

CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.

XX Sequence 14175 BP; 3111 A; 3818 C; 4137 G; 3109 T; 0 other;
SQ Query Match 3.8%; Score 43; DB 22; Length 14175;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 351 atgttgccagggttgctcaactctggtcgaagcaatcc 393
|||||
Db 9809 atgttgccagggttgctcaactctggtcgaagcaatcc 9851

RESULT 11
AAK78859/G
ID AAK78859 standard; DNA; 14175 BP.
XX
AC AAK78859;
XX
XX
DT 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33671.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytotatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
PN WO200157182-A2.
XX
XX 09-AUG-2001.
PF
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218220.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.
17-NOV-2000; 2000US-0249265.
17-NOV-2000; 2000US-0249297.
17-NOV-2000; 2000US-0249299.
17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis.

Disclosure: SEQ ID NO 33671; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins, and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAK82169
represent sequences used in the exemplification of the present invention.

SO Sequence 14175 BP; 3109 A; 4137 C; 3818 G; 3111 T; 0 other;

Query Match 3.8%; Score 43; DB 22; Length 14175;
Best Local Similarity :00.0%; Pred. No. 3.4e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 atgttgcacaggttggtctcaaaactcctgggtcgaagcaatcc 393
|||||
DB 4367 ATGTTGCCCAGGTGGTCTCAAACTCCTGGGCTCAAGCAATCC 4325

RESULT 12
AAK79678/c
ID AAK79678 standard; DNA; 23164 BP.
XX
AC AAK79678;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34490.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

```
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234323.
PR 21-SEP-2000; 2000US-0234374.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Disclosure; SEQ ID NO 34490; 3071pp + Sequence Listing: English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 23164 BP; 6164 A; 4977 C; 4888 G; 7135 T; 0 other;
XX
XX Query Match 3.8%; Score 43; DB 22; Length 23164;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-11;
XX Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 351 atgttcccgagttgtctcaaaactctggctcaagcaatcc 393
XX |
XX Db 3789 ATGTTGCCCCAGGTGGTCTCAAACTCTGGGCTCAAGCAATCC 3747
XX
XX RESULT 13
XX AAS29963/C
XX ID AAS29963 standard; DNA; 26995 BP.
XX
XX AC AAS29963;
XX
XX DT 21-NOV-2001 (first entry)
```

XX DE Human lung antigen genomic DNA #33.
XX KW Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;
KW KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW KW ophthalmological; vulnerrary; gene therapy; autoimmune disease; neoplasm;
KW KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW KW wound healing; skin aging; organ transplantation; food preservative;
KW KW tissue regeneration; anti-infertility; food additive.
XX OS Homo sapiens.
XX OS
PN WO200155303-A2.
XX PD
XX PF 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01301.
XX PF 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX respiratory disorders related to the lung including lung cancers and
XX also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO 227; 507pp; English.
XX
XX Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
XX the lung antigen polypeptides of the invention. Lung antigen polypeptides
XX and their associated polynucleotides are useful in the diagnosis,
XX treatment and prevention of various types of disorders in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
XX pathological condition can be determined by detecting the presence or
XX absence of a mutation in a lung antigen polynucleotide. The treatable
XX disorders include autoimmune diseases such as rheumatoid arthritis,
XX hyperproliferative disorders such as neoplasms of the breast or liver,
XX cardiovascular disorders such as cardiac arrest, cerebrovascular
XX disorders such as cerebral ischaemia, nervous system disorders such as
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi,
XX ocular disorders such as corneal infection, endocrine disorders such as
XX premature labour and infertility, gastrointestinal disorders such as
XX Crohn's disease, renal disorders such as glomerulonephritis and
XX respiratory disorders such as asthma and pleurisy. The polypeptides can
XX also be used to aid wound healing, to prevent skin aging due to sunburn,
XX to maintain organs before transplantation, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Query Match 3.8%; Score 43; DB 22; Length 26995;
XX Best Local Similarity 100.0%; Pred. No. 3.4e-11;
XX Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 346 tcacatgttgcagtggtgtctcaaacctctggctcaagc 388
XX |||||||
XX DB 5866 TCACCATGTTGCCAGGTGGTCTCAAACTCTGGCTCAAGC 5824
XX
XX RESULT 14
XX AAL16677
XX ID AAL16677 standard; CDNA; 4086 BP.
XX
XX AC AAL16677;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone kv10_8 nucleotide sequence SEQ ID NO:119.
XX
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antitumour; dermatological;
XX antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
XX antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX Homo sapiens.
OS
XX WO200009552-A1.
PN
XX 24-FEB-2000.
PD
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
PR 17-AUG-1998; 98US-0096815.
PR 04-SEP-1998; 98US-0099229.
PR 23-OCT-1998; 98US-0105368.
PR 08-JAN-1999; 99US-0115234.
PR 12-FEB-1999; 99US-0119931.
PR 18-FEB-1999; 99US-0120575.
PR 30-APR-1999; 99US-0132020.
PR 11-AUG-1999; 99US-0096622.
XX (GENY) GENETICS INST INC.
PA
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fechtel K;
XX WPI; 2000-205979/18.
XX P-PSDB; AAY94957.
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antitumour activity or tumor inhibition activity -
XX
XX Claim 128; Page 579-580; 641pp; English.
XX
XX AAL16618 to AAY94980 encode the human secreted proteins given in
XX AAY94988 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX markers on Southern gels; and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
XX diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
XX autoimmune inflammatory eye disease. The proteins can also be used to
XX treat allergic conditions, such as asthma. AAL16698 to AAL16774 represent
XX probes for the human secreted proteins from the present invention.
XX
XX Sequence 4086 BP; 1031 A; 942 C; 1014 G; 1099 T; 0 other;

Query Match 3.5%; Score 40; DB 21; Length 4086;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 346 tcaccatgtgtgccagggttggtctcaaaactcctcgtgggctca 395
|||||
Db 1982 tcaccatgtgtgccagggttggtctcaaaactcctcgtgggctca 2021

RESULT 15
AAK73037
ID AAK73037 standard; DNA; 12834 BP.
XX
AC AAK73037;
DT 06-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:27849.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytotstatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 13:46:30 ; Search time 1616.37 Seconds
(without alignments)
9527.537 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141

Sequence: 1 agataagcacagcagagaag.....aggataactgtgactccagg 1141

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hsc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	11.9	651	9	AL602439
2	97	8.5	738	10	BF208602
3	89	7.8	527	9	AL602677
4	89	7.8	999	10	BF676061
5	83	7.3	714	10	BF723255
6	59	5.2	442	10	H74001
7	55	4.8	412	10	H49484
8	55	4.8	648	12	AG100732
9	54	4.7	625	9	AL602554
10	47	4.1	288	9	AA382130
11	45	3.9	382	9	AA565270
12	45	3.9	395	9	AA713761
13	45	3.9	484	9	AA550845
14	45	3.9	931	10	BM008408
15	44	3.9	328	10	N53046
16	44	3.9	335	10	T96921
17	44	3.9	417	10	N73230

C 18	43	3.8	204	9	AW268077	AW268077	xf92d10.x
C 19	43	3.8	248	9	AA494274	AA494274	ng96b12.s
C 20	43	3.8	296	12	AQ089138	AQ089138	HS_2207_B
C 21	43	3.8	328	9	AA493880	AA493880	nh07a07.s
C 22	41	3.6	382	12	AQ147846	AQ147846	HS_3091_A
C 23	40	3.5	154	9	AA309434	AA309434	EST180303
C 24	40	3.5	293	9	AW961841	AW961841	EST373914
C 25	40	3.5	391	9	AW851810	AW851810	QVO-CT022
C 26	40	3.5	413	9	AL600265	AL600265	DKF2p3130
C 27	39	3.4	304	9	AI803767	AI803767	tp29h07.x
C 28	39	3.4	335	10	BI020761	BI020761	CM4-MT021
C 29	39	3.4	340	10	BI020795	BI020795	CM4-MT021
C 30	39	3.4	381	10	BF840682	BF840682	ILO-HT106
C 31	39	3.4	381	10	BF925361	BF925361	ILO-NT023
C 32	39	3.4	381	10	BF925367	BF925367	ILO-NT023
C 33	39	3.4	381	10	BF925371	BF925371	ILO-NT023
C 34	39	3.4	381	10	BF925374	BF925374	ILO-NT023
C 35	39	3.4	382	10	BF925369	BF925369	ILO-NT023
C 36	39	3.4	383	10	BF925368	BF925368	ILO-NT023
C 37	39	3.4	401	12	AQ481517	AQ481517	RPCI-11-2
C 38	39	3.4	416	10	BF896241	BF896241	ILO-MT021
C 39	39	3.4	416	10	BF896244	BF896244	ILO-MT021
C 40	39	3.4	418	10	BF896243	BF896243	ILO-MT021
C 41	39	3.4	617	12	B64146	B64146	CIT-HSP-202
C 42	39	3.4	639	9	AA316156	AA316156	EST187893
C 43	38	3.3	338	9	AA503501	AA503501	nh59h03.s
C 44	38	3.3	426	10	BF840283	BF840283	RC4-HT098
C 45	38	3.3	443	9	AA720734	AA720734	nv93d04.r

ALIGNMENTS

RESULT 1

AL602439 651 bp mRNA linear EST 14-AUG-2001
LOCUS DKFZp686L1213_r1 686 (synonym: hicc3) Homo sapiens CDNA clone
DEFINITION DKFZp686L1213 5', mRNA sequence.
ACCESSION AL602439
VERSION AL602439.1 GI:15165945
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS Ansonge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Ansonge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansonge W
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

No sequence available.

This clone (DKFZp686L1213) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de.

FEATURES

Location/Qualifiers
1..651
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686L1213"
/clone.lib="686 (synonym: hicc3)"
/tissue.type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplEx2; Site_1: SflIA; Site_2: SflIB; cDNA-collection"

By

KEYWORDS SOURCE ORGANISM

EST. human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 999)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

REFERENCE AUTHORS TITLE JOURNAL COMMENT

Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M1066 row: h column: 10
High quality sequence stop: 562.
Location/Qualifiers
1. 999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4248249"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfII (ggcgctcgcc); Site_2: SfiI (ggccataggcc
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGCGCATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

FEATURES

source

BASE COUNT 273 a 269 c 241 g 216 t

ORIGIN
Query Match 7.8%; Score 89; DB 10; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.1e-32;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1053 agacggtcgctccacatccaggtctttgtcctcgtcgtcgtcttcttccac 1112
|||||
Db 2 AGACGGTGGCTCCACATCCAGGTCTTTGTGCTCCTGCTGCTCTTTCCAC 61
QY 1113 gcatttccagataactgtgactccagg 1141
|||||
Db 62 GCATTTTCCAGGTAACCTGACTCCAGG 90
RESULT 5
BG723255
LOCUS
DEFINITION
607690875F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4823059 5',
mRNA sequence.
ACCESSION
BG723255
VERSION
BG723255.1 GI:14002442
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 114)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10732 row: f column: 20
High quality sequence stop: 711.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4823059"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

FEATURES

source

BASE COUNT 211 a 167 c 147 g 187 t
ORIGIN
Query Match 7.3%; Score 83; DB 10; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.6e-29;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1059 gtcgctccacatccaggtctttgtcctcgtcgtcgtcttcttccagcatt 1118
|||||
Db 6 GGTGGCTCCATCCAGGTCTTTGTGCTCCTGCTGCTGCTTCTTCCAGCATTT 65
QY 1119 tcaggataactgtgactccagg 1141
|||||
Db 66 TCCAGATAACTGTGACTCCAGG 88
RESULT 6
W74001
LOCUS
DEFINITION
W74001 442 bp mRNA linear EST 20-JUN-1996
IMAGE:345228 5' similar to gb:U04313 MASPIN PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION
W74001
VERSION
W74001.1 GI:1384551
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 442)
REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 182.
Location/Qualifiers

FEATURES

source

```

1. .442
/organism="Homo sapiens"
/db_xref="GDB:1270603"
/db_xref="taxon:9606"
/clone="IMAGE:345228"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; lstr
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAAGTGGAGCGCCGACATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
BASE COUNT      107 a      83 g      123 t      13 others
ORIGIN

Query Match      5.2%; Score 59; DB 10; Length 442;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1083 gtgtctctgtgctgttcttctccacgcgcatctttccaggataactgtgactccagg 1141
|||||
DB 45 GTGCTCCTCGTTCCTGCTCTCTTTCCACGCAATTTTCCAGGATAACTGTGACTCCAGG 103

RESULT 7
LOCUS      H49484      412 bp      mRNA      linear      EST 22-JAN-1999
DEFINITION      yq20d10.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:274242 3' similar to contains Alu repetitive element; contains
THR repetitive element ;, mRNA sequence.
ACCESSION      H49484
VERSION      H49484.1 GI:989325
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 412)
AUTHORS      Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston,
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1537
High quality sequence stops: 301
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1537 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 301.
Location/Qualifiers
1. .412
/organism="Homo sapiens"
/db_xref="GDB:3798193"
/db_xref="taxon:9606"

FEATURES
source

```

```

/clone="IMAGE:274242"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      92 a      82 g      106 t      4 others
ORIGIN

Query Match      4.8%; Score 55; DB 10; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 tttttgtagagtggtctcaccatgttgcagggttggtctcaactccctggg 381
|||||
DB 172 TTTTGTAGAGTGGGTCTCACCATGTGCCAGGTTGGTCTCAACTCTCTGGG 226

RESULT 8
LOCUS      AG100732      648 bp      DNA      linear      GSS 03-NOV-2001
DEFINITION      Pan troglodytes DNA, clone: PTB-103G10.R, genomic survey sequence.
ACCESSION      AG100732
VERSION      AG100732.1 GI:16721249
KEYWORDS      GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
BAC Library clone:PTB-103G10.R.
ORGANISM      Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      1 (sites)
AUTHORS      Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 648)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpesgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1. .648
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-103G10.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      192 a      138 c      156 g      161 t
ORIGIN

```

Query Match 4.8%; Score 55; DB 12; Length 648;
 Best Local Similarity 100.0%; Pred. No. 1.1e-15;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 tttttgtagagatggggtccaccatgttgcagagttggtctcaaacctcctggg 381
 |||||
 Db 562 TTTTGTAGAGATGGGTCACCATGTGTGCCAGGTTGGTCAAACTCCTGGG 508

RESULT 9
 AL602554 625 bp mRNA linear EST 14-AUG-2001
 DEFINITION DKFZp686F1514_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
 DKFZp686F1514 5', mRNA sequence.
 AL602554
 VERSION AL602554
 KEYWORDS EST.
 SOURCE AL602554.1 GI:15166060
 ORGANISM human.

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 625)
 Wambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.
 EST (Wambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.)
 Unpublished (1999)
 CONTACT: Wambutt R
 MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOWA (Berlin/Germany) within the CDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZp686F1514) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 source
 1..625
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp686F1514"
 /clone_lib="686 (synonym: hlcc3)"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"
 BASE COUNT 180 a 149 c 122 g 174 t
 ORIGIN

Query Match 4.7%; Score 54; DB 9; Length 625;
 Best Local Similarity 100.0%; Pred. No. 3.5e-15;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1088 cctcgctgcctgtcttccagcatttccaggataactgtgactccagg 1141
 |||||
 Db 32 CCTCGCTGCCTGCTCCCTTTCCAGCATTTCCAGGATACTGTGACTCCAGG 85

RESULT 10
 AA382130 288 bp mRNA linear EST 21-APR-1997
 LOCUS EST95313 Activated T-cells II Homo sapiens cDNA 5' end similar to
 DEFINITION EST containing Alu repeat, mRNA sequence.
 AA382130
 ACCESSION AA382130.1 GI:2034469
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 288)
 Adams M.D., Kerlavage, A.R., Fleischmann, R.D., Fulder, R.A., Bult
 C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S. Jr., Kelley, J.M.,
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
 Moreno-Pallanques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dinke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
 Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
 M.R., Rosen, C.A., Hasetline, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.

TITLE
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL MEDLINE COMMENT

96026280
 Other_ESTs: EST95312
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018599056
 Fax: 3018599423

Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse

FEATURES
 source
 Location/Qualifiers
 1..288
 /organism="Homo sapiens"
 /db_xref="ATCC (Inhost):186511"
 /db_xref="taxon:9606"
 /clone_lib="Activated T-cells II"
 /cell_type="T-lymphocyte"
 /dev_stage="adult"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 79 a 64 c 44 g 100 t 1 others
 ORIGIN

Query Match 4.1%; Score 47; DB 9; Length 288;
 Best Local Similarity 100.0%; Pred. No. 8.8e-12;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 tcaccatgttgcaggttggtctcaaacctcctgggtcgaagcaatc 392
 |||||
 Db 15 TCACCATGTGTCAGGTGGTCTCAAACTCCTGGGCTCAAGCAATC 61

RESULT 11
 AA565270/c 382 bp mRNA linear EST 08-SEP-1997
 LOCUS nk53b11.s1 NCI-CGAP_P7 Homo sapiens cDNA clone IMAGE:1017213
 DEFINITION similar to contains Alu repetitive element,, mRNA sequence.
 AA565270
 ACCESSION AA565270.1 GI:2336909
 VERSION
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 382)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 533 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 341.
Location/Qualifiers

FEATURES
source

1. .382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1017213"
/clone_lib="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH108"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
prostate intrapitthelial neoplasia (high-grade), cDNA
made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp."

BASE COUNT 112 a 77 c 92 g 101 t

ORIGIN

Query Match 3.9%; Score 45; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 8.3e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 ccattgtgccagggtgtctcaaacctcctgggtcgaagcaatcc 393
|||||
Db 222 CCATGTTGCCAGGTGCTCCTCAACTCTCTGGGCTCAAGCAATCC 178.

RESULT 12
AA713761

LOCUS
DEFINITION
nv82e02 s1 NCI_CGAP_Pr4 395 bp mRNA linear EST 18-FEB-1998
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION
AA713761

VERSION
AA713761.1 GI:2726035

KEYWORDS
EST.

SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 395)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
Ph.D. student, Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 458 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers

FEATURES

source

1. .395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1236314"
/clone_lib="NCI_CGAP_Pr4"
/sex="female"
/tissue_type="normal ductal tissue"
/lab_host="DH108"
/note="Organ: breast; Vector: pAMP10; mRNA made from
normal breast ductal tissue, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Reference:
Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 100 a 99 c 88 g 108 t

ORIGIN

Query Match 3.9%; Score 45; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 8.4e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 coatttgcagggtgtctcaaacctcctgggtcgaagcaatcc 393
|||||
Db 206 CCATGTTGCCAGGTGCTCCTCAACTCTCTGGGCTCAAGCAATCC 250

RESULT 13
AA550845

LOCUS
DEFINITION
nj34e10.s1 NCI_CGAP_AAL Homo sapiens cDNA clone IMAGE:994410 3'
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION
AA550845

VERSION
AA550845.1 GI:2321097

KEYWORDS
EST.

SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 484)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 434.
Location/Qualifiers

FEATURES
source

1. .484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="994410"
/clone_lib="NCI_CGAP_AAL"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: adrenal gland; Vector: Bluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Two pooled bulk adrenal adenomas.
5' adaptor sequence: 5' GAATTCGGCAGG 3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3' Average insert
size: 1.6 kb."
BASE COUNT 130 a 104 c 99 g 151 t

ORIGIN

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 328)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chitapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 403 Std Error: 0.00
Seq primer: ml3 -40 forward
High quality sequence stop: 378.
Location/Qualifiers
1..328
/organism="Homo sapiens"
/db_xref="GDB:3795723"
/db_xref="taxon:9606"
/clone_lib="IMAGE:246477"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACGTGAAGAATTAATTAAGATCTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 69 a 73 c 70 g 116 t
ORIGIN

Query Match 3.9%; Score 44; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 350 catgtgtccaggttgctcacaactcctgggtcgaagcaatcc 393
|||||
Db 19 CATGTTGCCAGGTGGTCTCAAACTCCTGGGTCAAGCAATCC 62

Search completed: September 24, 2002, 16:11:34
Job time: 8704 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 10:15:05 : Search time 3102.12 seconds
(without alignments)
7697.052 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagcagagaag.....aggataactgtgactccagg 1141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_to.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

1	1129.8	99.0	170269	9	AC090307	AC090307 Homo sapi
2	1129.8	99.0	173131	2	AC036176	AC036176 Homo sapi
3	1129.8	99.0	173836	2	AP001404	AP001404 Homo sapi
4	745.4	65.3	12426	6	AX251236	AX251236 Sequence
5	723.4	63.4	12426	6	AX251235	AX251235 Sequence
6	612	53.6	214212	2	AC015793	AC015793 Homo sapi
7	437.6	38.4	214212	2	AC015793	AC015793 Homo sapi
8	203.4	17.8	44255	9	AL356309	AL356309 Human DNA
9	199.4	17.5	84577	9	AC079927	AC079927 Homo sapi
10	195	17.1	2343	9	AK023951	AK023951 Homo sapi
11	195	17.1	71872	9	AL133390	AL133390 Human DNA
12	195	17.1	147259	2	AC019248	AC019248 Homo sapi
13	195	17.1	184703	9	AL365215	AL365215 Human DNA
14	194	17.0	164126	2	AC090104	AC090104 Homo sapi
15	194	17.0	169669	2	AC041005	AC041005 Homo sapi
16	194	17.0	190689	2	AC091185	AC091185 Homo sapi
17	193.4	17.0	158806	2	AL591132	AL591132 Homo sapi
18	193.4	17.0	168834	2	AL355583	AL355583 Homo sapi
19	193.4	17.0	200882	2	AL591181	AL591181 Homo sapi
20	193.4	17.0	214180	2	AL360082	AL360082 Homo sapi
21	193.4	17.0	238379	2	AL590986	AL590986 Homo sapi
22	193.4	17.0	300000	9	AL645813	AL645813 Human DNA
23	192.8	16.9	161296	2	AC012308	AC012308 Homo sapi
24	192.6	16.9	196533	9	CNS007VI	AL096870 Human chr
25	191.6	16.8	148555	2	AP004711	AP004711 Homo sapi
26	191.6	16.8	155614	2	AC012620	AC012620 Homo sapi
27	191.6	16.8	160887	2	AC021498	AC021498 Homo sapi
28	191.6	16.8	225037	2	AC091825	AC091825 Homo sapi
29	191.4	16.8	162554	9	AC079354	AC079354 Homo sapi
30	190.8	16.7	102837	9	HS52202	295329 Human DNA s
31	190.8	16.7	183245	2	AC099397	AC099397 Homo sapi
32	190.6	16.7	199179	2	AC024619	AC024619 Homo sapi
33	190.4	16.7	109902	9	HSDJ46C2	AL132775 Human DNA
34	190.4	16.7	178964	2	AC012468	AC012468 Homo sapi
35	190.2	16.7	150477	2	AC024441	AC024441 Homo sapi
36	190.2	16.7	159265	2	AC068307	AL157942 Homo sapi
37	190.2	16.7	168759	2	AC084871	AC084871 Homo sapi
38	190.2	16.7	203918	2	AC084871	AL357131 Homo sapi
39	190	16.7	110000	2	AL357131_0	AL357131 Homo sapi
40	190	16.7	157044	2	AC079849	AC079849 Homo sapi
41	190	16.7	180533	2	AL589943	AL589943 Homo sapi
42	190	16.7	187052	2	AC078926	AC078926 Homo sapi
43	190	16.7	188923	9	AC020612	AC020612 Homo sapi
44	190	16.7	215867	2	AC084037	AC084037 Homo sapi
45	190	16.7	257632	2	AC009246	AC009246 Homo sapi

ALIGNMENTS

RESULT 1	AC090307	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
AC090307	LOCUS	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.				
DEFINITION	AC090307					
ACCESSION	AC090307					
VERSION	HTG					
KEYWORDS	human.					
SOURCE	ORGANISM					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 170269)					
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.					
TITLE	Homo sapiens chromosome 18, clone RP11-851B10					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 170269)					
AUTHORS	Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,					

[illegible]

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 173836)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Homo sapiens 173,836 genomic DNA of 18q22
 Published Only in Database (2000) In press
 2 (bases 1 to 173836)
 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P., Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
 Direct Submission
 Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-42-778-9923, Fax: 81-42-778-9924)
 On Jul 14, 2000 this sequence version replaced gi:8117315.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center (GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft18
 Center clone name: RP11-851B10
 ----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 171265 bases at least Q40
 Consensus quality: 172100 bases at least Q30
 Consensus quality: 172469 bases at least Q20
 Insert size: 172836; sum-of-contigs
 Quality coverage: 10.50x in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved
 1 68896 contig of 68896 bp in length
 68997 95899 contig of 26903 bp in length
 96000 113268 contig of 23269 bp in length
 119369 141739 contig of 23271 bp in length
 141840 152655 contig of 10816 bp in length
 152756 159466 contig of 6711 bp in length
 159567 163465 contig of 3899 bp in length
 163566 166433 contig of 2868 bp in length
 166534 170112 contig of 3579 bp in length
 170213 172095 contig of 1883 bp in length
 172196 173836 contig of 1641 bp in length.
 * NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
 1 68896: contig of 68896 bp in length
 68997 68996: gap of 100 bp
 95899 95899: contig of 26903 bp in length
 96000 95999: gap of 100 bp
 119369 119268: contig of 23269 bp in length
 141840 141739: contig of 23271 bp in length
 152756 152655: contig of 10816 bp in length
 159466 159466: gap of 100 bp
 159567 159566: gap of 100 bp

* 159567 153465: contig of 3899 bp in length
 * 163466 163565: gap of 100 bp
 * 163566 166433: contig of 2868 bp in length
 * 166434 166533: gap of 100 bp
 * 166534 170112: contig of 3579 bp in length
 * 170113 170212: gap of 100 bp
 * 170213 172095: contig of 1883 bp in length
 * 172096 172195: gap of 100 bp
 * 172196 173836: contig of 1641 bp in length.
 FEATURES
 Location/Qualifiers
 1..173836
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q22"
 /clone="RP11-851B10"
 1..68896
 /note="assembly_fragment"
 68997..95899
 /note="assembly_fragment clone_end:T7 vector_side:right"
 96000..119268
 /note="assembly_fragment"
 119369..141739
 /note="assembly_fragment"
 141840..152655
 /note="assembly_fragment"
 152756..159466
 /note="assembly_fragment"
 159567..163465
 /note="assembly_fragment"
 163566..166433
 /note="assembly_fragment clone_end:SP6 vector_side:left"
 166534..170112
 /note="assembly_fragment"
 170213..172095
 /note="assembly_fragment"
 172196..173836
 /note="assembly_fragment"
 BASE COUNT 53047 a 34410 c 33807 g 51570 t 1002 others
 ORIGIN
 Query Match 99.0%; Score 1129.8; DB 2; Length 173836;
 Best Local Similarity 99.4%; Pred. No. 8.1e-309;
 Matches 1134; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 agataagcacagcagagaagcaaccagctccgtttcaggtctcttctcaggtgattcg 60
 Db 150724 AGATAAGCACAGCAGAGAAGCAACCAAGCTCCGTTTCAGGTCTCTCTTCTGAGACTGATTCG 150665
 QY 61 gctggaaggagtagtcccacccaaatgaagaagctgtgggaagacagagacaagaac 120
 Db 150664 GCTGGAAGGAGTAGGTCCCGCAATGAAGAAGCTGTGGGAAGACAGAGGACAAGAAC 150605
 QY 121 aggtccacagaagagatttcagagcagagctgcgtactctcttcttcttcttcttctt 180
 Db 150604 AGGCTCCACGAAGAGATTTTCAGAGCAGAGCTCGGTACTCTCTTTCTTTTCTTTT 150545
 QY 181 gctctgtaccacaggtgaaagtacagtggttagctcaagctcaggtcaggtttgacctc 240
 Db 150544 GCTCTGTACCCAGGCTGAAGTACAGTGTAGTCTACCTCAGCTCAGCTGACGCTT 150485
 QY 241 ccaggtcgaagtagctctctctcagcttcccaagtaactgggaccacagcagcatgc 300
 Db 150484 CCAGGCTCAAGTAGATCTCTCGTCTCAGCTTTCCAAAGTAAGTGGGACCAAGCATGCAT 150425
 QY 301 caccacgttaggcattgttttacattttttagagatggggtctccacattttgccca 360
 Db 150424 CACCACACTAGGCTATTGTTTACATTTTGTAGAGATGGGCTCTCACCATGTTGCCCA 150365
 QY 361 ggttggtctcaaaactcctgggctcaagcaatccgcgtcaacgtcaacccccaaatgctgg 420
 Db 150364 GGTGTGCTCAAACTCCTGGGCTCAAGCAATCGGCTCAGCTCAACCTCCCAAAATGCTGG 150305

/db_xref="taxon:32630"	
/note="chemically treated genomic DNA (Homo sapiens)"	
BASE COUNT	3791 a 137 c 2517 g 5981 t
ORIGIN	
Query Match 65.3%; Score 745.4; DB 6; Length 12426;	
Best Local Similarity 78.4%; Pred. No. 3.7e-200;	
Matches 893; Conservative 0; Mismatches 246; Indels 0; Gaps 0;	
QY 1	agataagcagcagcagagagaacacagctcgcgtttccaggtctcttccctgagctgattcg 60
DB 8506	AAATAAACACACAAAAAACAACAACACTCCGTTTCAAAATCTTTCTCCFAAACATAATTCG 8447
QY 61	gctgaaagggagtaggtccaccacaaatgaagaagctgtggaagacagagagacaagaac 120
DB 8446	ACTAAAAAATAAATATCCGCCAATTAATAAAACATATAAAAAAACAACAAAAACA 8387
QY 121	aggctccacgaagagatttcagagcagagctgctactcctcttcttcttcttcttctt 180
DB 8386	AAATCCACGAAAAAATTTCAAAACAAACTACGTACTCTCTTTTCTTTTATTTCTTTT 8327
QY 181	gctctgtaacccaggtgaagtacagtggttagctcagcgtcactcagcttgaacctc 240
DB 8326	ACTCTATCACCACAACTAAATACAAATAATTAACACAGCTCAGCTCACTCAACTTAA 8267
QY 241	ccaggtcgaagtgcctctcgtcagcttccaaagtaactgggaccacagggcatgcat 300
DB 8266	CCAAACTCAAAATAATCTCTCGTCTCAACTTTTCCAAATAACTTAAACACCAACATAC 8207
QY 301	caccacgtcagctattgtttacatttttttagagatggggtctcaccatgttgcaca 360
DB 8206	CACCACACTAAACTATATTTTACATTTTTTATAAAAAATAAATCTACCATATTAACCA 8147
QY 361	ggctgtctcaactcctggctcaagcaatccgcgtccacgtccaaactccccaaatgtcg 420
DB 8146	AATTAATCTCAAACTCTTAACCTCAAAACAATCCGCTCAGCTCAACCTCCCAAACTAA 8087
QY 421	gattacagcgtgagccacccggcggcggcgtgagtaactcctcaatcacaggaatttcaa 480
DB 8086	AATTAACAACTAAACACCGCGCCAAACCTAAATTAATCTTAATCAACAAATTTTAAAA 8027
QY 481	gaactctcgtccaccacattaaacaatatctccacaaatttggtagtaaatatttg 540
DB 8026	AAACCTTCTTACGCCACCCATTAACAATAATCTCTCTACCAATTTAATAATAATTTTA 7967
QY 541	ctaatagtacctaatatttttagtgagcactgtgtttatacatatccatctctctttt 600
DB 7966	CTAATAATACCTAATTTTAAATAAACACTATATTTATATATATATATATATCTCTCTT 7907
QY 601	ttgattgtcttctgttttaagggcagctaccctctcttggcatctagcagaatgagctgc 660
DB 7906	TTAATATATCTTCTATTTTAAATAAACACTACCTCTCTTAACATCTAACAAAAATAACTAC 7847
QY 661	tgcaagttcacaaaaagaatggagatcagagctacttttggccaccacagtgctgtgag 720
DB 7846	TACAAATTTACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7787
QY 721	aaatttgtagtgtaactatcatcacattactttttatttttcacgaatatattcaccttc 780
DB 7786	AAATTTATAATATTTACTATCATCATCATCATCTTTTATTTTATTCAGAAATATTTTCCCTTCC 7727
QY 781	ggctcgtggtggcggagagattgcgtgacgcatgtctgtatgcatgtatgcatgtaactcac 840
DB 7726	GATCTAGTAAACCGAAAAAATATACCGTAGCATATCTATAGTATACATATAAATACTAC 7667
QY 841	agcccttctcgtcccgacaatgttggaggccttttggagactgtgacagacaacagcaact 900
DB 7666	AACCCCTTCTTACCCGAACATATTAATAAAACCTTTTAAAAAACTATACAAAAACAATAA 7607
QY 901	tcagcctgaatcatctcttccaattgtggacaagctcccaagcgttggtagtaggagagg 960
DB 7606	TCAACCTAAATCATTTCTTCTTCAATTAATAAAACACTACCAAAAACTTAAATAAAAAA 7547

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 3, 2000 this sequence version replaced gi:5446894.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: 1117_D_15

 * NOTE: This record contains 240 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 * 696 795: contig of 695 bp in length
 * 796 1513: contig of 718 bp in length
 * 1514 1613: gap of 100 bp
 * 1614 2337: contig of 724 bp in length
 * 2338 2437: gap of 100 bp
 * 2438 3161: contig of 724 bp in length
 * 3162 3261: gap of 100 bp
 * 3262 3966: contig of 705 bp in length
 * 3967 4066: gap of 100 bp
 * 4067 4772: contig of 706 bp in length
 * 4773 4872: gap of 100 bp
 * 4873 5564: contig of 692 bp in length
 * 5565 5664: gap of 100 bp
 * 5665 6397: contig of 733 bp in length
 * 6398 6497: gap of 100 bp
 * 6498 7202: contig of 705 bp in length
 * 7203 7302: gap of 100 bp
 * 7303 8028: contig of 726 bp in length
 * 8029 8128: gap of 100 bp
 * 8129 8803: contig of 675 bp in length
 * 8804 8903: gap of 100 bp
 * 8904 9613: contig of 710 bp in length
 * 9614 9713: gap of 100 bp
 * 9714 10409: contig of 696 bp in length
 * 10410 10509: gap of 100 bp
 * 10510 11237: contig of 728 bp in length
 * 11238 11337: gap of 100 bp
 * 11338 12043: contig of 706 bp in length
 * 12044 12143: gap of 100 bp
 * 12144 12867: contig of 724 bp in length
 * 12868 12967: gap of 100 bp
 * 12968 13660: contig of 693 bp in length
 * 13661 13760: gap of 100 bp
 * 13761 14489: contig of 729 bp in length
 * 14490 14589: gap of 100 bp
 * 14590 15314: contig of 725 bp in length
 * 15315 15414: gap of 100 bp
 * 15415 16129: contig of 715 bp in length
 * 16130 16229: gap of 100 bp
 * 16230 16946: contig of 717 bp in length
 * 16947 17046: gap of 100 bp
 * 17047 17765: contig of 723 bp in length
 * 17770 17869: gap of 100 bp
 * 17870 18574: contig of 705 bp in length
 * 18575 18674: gap of 100 bp
 * 18675 19379: contig of 705 bp in length
 * 19380 19479: gap of 100 bp
 * 19480 20186: contig of 707 bp in length
 * 20187 20286: gap of 100 bp
 * 20287 20980: contig of 694 bp in length
 * 20981 21080: gap of 100 bp
 * 21081 21779: contig of 699 bp in length
 * 21780 21879: gap of 100 bp
 * 21880 22594: contig of 715 bp in length
 * 22595 22694: gap of 100 bp
 * 22695 23399: contig of 705 bp in length
 * 23400 23499: gap of 100 bp
 * 23500 24217: contig of 718 bp in length
 * 24218 24317: gap of 100 bp
 * 24318 25026: contig of 709 bp in length
 * 25027 25126: gap of 100 bp
 * 25127 25836: contig of 710 bp in length
 * 25837 25936: gap of 100 bp
 * 25937 26637: contig of 701 bp in length
 * 26638 26737: gap of 100 bp
 * 26738 27446: contig of 709 bp in length
 * 27447 27546: gap of 100 bp
 * 27547 28274: contig of 728 bp in length
 * 28275 28374: gap of 100 bp
 * 28375 29095: contig of 721 bp in length
 * 29096 29195: gap of 100 bp
 * 29196 29917: contig of 722 bp in length
 * 29918 30017: gap of 100 bp
 * 30018 30724: contig of 707 bp in length
 * 30725 30824: gap of 100 bp
 * 30825 31529: contig of 705 bp in length
 * 31530 31629: gap of 100 bp
 * 31630 32324: contig of 695 bp in length
 * 32325 32424: gap of 100 bp
 * 32425 33116: contig of 692 bp in length
 * 33117 33216: gap of 100 bp
 * 33217 33936: contig of 720 bp in length
 * 33937 34036: gap of 100 bp
 * 34037 34750: contig of 714 bp in length
 * 34751 34850: gap of 100 bp
 * 34851 35578: contig of 728 bp in length
 * 35579 35678: gap of 100 bp
 * 35679 36393: contig of 715 bp in length
 * 36394 36493: gap of 100 bp
 * 36494 37203: contig of 710 bp in length
 * 37204 37303: gap of 100 bp
 * 37304 37965: contig of 662 bp in length
 * 37966 38065: gap of 100 bp
 * 38066 38781: contig of 716 bp in length
 * 38782 38881: gap of 100 bp
 * 38882 39590: contig of 709 bp in length
 * 39591 39690: gap of 100 bp
 * 39691 40403: contig of 713 bp in length
 * 40404 40503: gap of 100 bp
 * 40504 41238: contig of 735 bp in length
 * 41239 41338: gap of 100 bp
 * 41339 42064: contig of 726 bp in length
 * 42065 42164: gap of 100 bp
 * 42165 42880: contig of 716 bp in length
 * 42881 42980: gap of 100 bp
 * 42981 43687: contig of 707 bp in length
 * 43688 43787: gap of 100 bp
 * 43788 44504: contig of 717 bp in length
 * 44505 44604: gap of 100 bp

TITLE
 JOURNAL

COMMENT

Db 168125 GAAACTTCTCGCCACCCATAAACAATATCTCTACCAATTTGGTAGTAATAATTTTG 168066
 Qy 541 ctaatagcaccaatatttttagtaggactgtgtttatcacatataccatccctctttt 600
 |||||
 Db 168065 CTAATAGTACCTAATTTTAGTAGGACACTGTGTTTATACATATATCATTCCTCTTTT 168006
 Qy 601 ttgattgtcttctgttttaatggcagctacacctctcttggcattagcagaatgagctgc 660
 |||||
 Db 168005 TTGATTGTCTTCTGTGTTTAATGAATTCACGCACATGGACGTTTCAGCTGCTCCTNNNNNN 167946
 Qy 661 tgcagtttacacaaaagaatggagatcagagtagtactttttgtgcacacaagctgtctgag 720
 |||||
 Db 167945 NNN 167886
 Qy 721 aaattgtagtgttactatcatcacattacttttttttttttttttttttttttttttt 769
 |||||
 Db 167885 NNN 167837
 |||||

RESULT 7
 AC015793 214212 bp DNA linear HTG 13-JUL-2000
 LOCUS Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
 DEFINITION SEQUENCE SAMPLING.
 AC015793
 AC015793.2 GI:7144769
 VERSION HTG: HTGS_PHASE0.
 KEYWORDS human.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 214212)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 JOURNAL Homo sapiens chromosome 18, clone RP11-1117D15
 TITLE Unpublished
 REFERENCE 2 (bases 1 to 214212)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgaiter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferrelle,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehotsky,J., Lisu,C., Locke,K., MacDonald,P., Marquis,N.,
 McEwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 DIRECT SUBMISSION
 TITLE Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
 JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Mar 3, 2000 this sequence version replaced gi:6446894.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L1590
 Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

*
* 1
* 696 795: gap of 100 bp
* 796 1513: contig of 695 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12968 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14589: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp
* 20287 20980: contig of 694 bp in length
* 20981 21080: gap of 100 bp
* 21081 21779: contig of 699 bp in length
* 21780 21879: gap of 100 bp
* 21880 22594: contig of 715 bp in length
* 22595 22694: gap of 100 bp
* 22695 23398: contig of 705 bp in length
* 23400 23499: gap of 100 bp
* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length
*
* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp
* 34851 35578: contig of 728 bp in length
* 35579 35678: gap of 100 bp
* 35679 36393: contig of 715 bp in length
* 36394 36493: gap of 100 bp
* 36494 37203: contig of 710 bp in length
* 37204 37303: gap of 100 bp
* 37304 37965: contig of 662 bp in length
* 37966 38065: gap of 100 bp
* 38066 38781: contig of 716 bp in length
* 38782 38881: gap of 100 bp
* 38882 39590: contig of 709 bp in length
* 39591 39690: gap of 100 bp
* 39691 40403: contig of 713 bp in length
* 40404 40503: gap of 100 bp
* 40504 41238: contig of 735 bp in length
* 41239 41338: gap of 100 bp
* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length
* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 54554 55184: gap of 100 bp
* 55085 55185: contig of 788 bp in length
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp

RESULT 9

AC079927 84577 bp DNA linear PRI 30-JAN-2002
Homo sapiens BAC clone RP11-702A23 from 4, complete sequence.

AC079927
AC079927 5 GI:18071472

HTG.

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 84577)

Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998)

99063792

2 (bases 1 to 84577)

Cedroni, M., Abbott, A. and Boyer, E.

The sequence of Homo sapiens BAC clone RP11-702A23

Unpublished (2001)

3 (bases 1 to 84577)

Waterston, R.H.

Direct Submission

Submitted (17-SEP-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

4 (bases 1 to 84577)

Waterston, R.H.

Direct Submission

Submitted (05-JAN-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 84577)

Waterston, R.

Direct Submission

Submitted (30-JAN-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Jan 5, 2002 this sequence version replaced gi:16924160.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.edu

----- Summary Statistics

----- Center project name: H_NH0702A23

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from

Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-757K22, 2000 bp overlap; the clone sequenced to the right is RP11-19F13. Actual start of this clone is at base position 110812 of RP11-757K22; actual end is at base position 84577 of RP11-702A23.

Two transposons occurred in the growth of the clone that have been omitted from the finished sequence.

FEATURES	Source
repeat_region	1..84577
repeat_region	Location/Qualifiers
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="4"
repeat_region	/map="4"
repeat_region	/clone="RP11-702A23"
repeat_region	/clone_lib="RPC1-11"
repeat_region	1..206
repeat_region	/rpt_family="Alu"
repeat_region	256..528
repeat_region	/rpt_family="Alu"
repeat_region	697..723
repeat_region	/rpt_family="(TTTC)n"
repeat_region	713..846
repeat_region	/rpt_family="Alu"
repeat_region	1407..1747
repeat_region	/rpt_family="MER1_type"
repeat_region	1799..2262
repeat_region	/rpt_family="L2"
repeat_region	2265..2290
repeat_region	/rpt_family="(TTTAA)n"
repeat_region	2266..2574
repeat_region	/rpt_family="Alu"
repeat_region	2599..2719
repeat_region	/rpt_family="Alu"
repeat_region	2891..2996
repeat_region	/rpt_family="MIR"
repeat_region	3255..3607
repeat_region	/rpt_family="MER2_type"
repeat_region	3620..3713
repeat_region	/rpt_family="MER53"
repeat_region	3637..3719
repeat_region	/rpt_family="MER53"
repeat_region	4055..4083
repeat_region	/rpt_family="AT_rich"
repeat_region	4056..4358
repeat_region	/rpt_family="Alu"
repeat_region	4698..4721
repeat_region	/rpt_family="AT_rich"
repeat_region	5095..5462
repeat_region	/rpt_family="MaLR"
repeat_region	6970..7010
repeat_region	/rpt_family="(TTC)n"
repeat_region	6985..7294
repeat_region	/rpt_family="Alu"
repeat_region	7703..7992
repeat_region	/rpt_family="Alu"
repeat_region	8526..8559
repeat_region	/rpt_family="(T)n"
repeat_region	8532..8840
repeat_region	/rpt_family="Alu"
repeat_region	9438..9458
repeat_region	/rpt_family="AT_rich"
repeat_region	9440..9738
repeat_region	/rpt_family="Alu"
repeat_region	9744..9886
repeat_region	/rpt_family="MIR"
repeat_region	10404..10507
repeat_region	/rpt_family="L2"

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:5730909.
All repeats were identified using RepeatMasker;
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1869
Center Clone name: 115_G_24
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141535 bases at least Q40
Consensus quality: 144279 bases at least Q30
Consensus quality: 145529 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 146459; sum-of-contigs
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2950: contig of 2950 bp in length
* 2951 3050: gap of 100 bp
* 3051 6314: contig of 3264 bp in length
* 6315 6414: gap of 100 bp
* 6415 11824: contig of 5410 bp in length
* 11825 11924: gap of 100 bp
* 11925 20145: contig of 8221 bp in length
* 20146 20245: gap of 100 bp
* 20246 30884: contig of 10639 bp in length
* 30885 30984: gap of 100 bp
* 30985 44984: contig of 14000 bp in length
* 44985 45084: gap of 100 bp
* 45085 77877: contig of 32793 bp in length
* 77878 77977: gap of 100 bp
* 77978 112578: contig of 34601 bp in length
* 112579 112679: gap of 100 bp
* 112679 147259: contig of 34581 bp in length.
Location/Qualifiers
1. 147259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-115624"
/clone_lib="RPC1-II Human Male BAC"
1. 2950
/note="assembly_fragment"
3051. 6314
/note="assembly_fragment"
6415. 11824
/note="assembly_fragment"
11925. 20145
/note="assembly_fragment"
20246. 30884
/note="assembly_fragment"
30985. 44984
/note="assembly_fragment"
clone_end:17
vector_side:right
45085. 77877
/note="assembly_fragment"
77978. 112578

COMMENT
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:5730909.
All repeats were identified using RepeatMasker;
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1869
Center Clone name: 115_G_24
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 141535 bases at least Q40
Consensus quality: 144279 bases at least Q30
Consensus quality: 145529 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 146459; sum-of-contigs
Quality coverage: 5.3 in Q20 bases; agarose-fp
Quality coverage: 5.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2950: contig of 2950 bp in length
* 2951 3050: gap of 100 bp
* 3051 6314: contig of 3264 bp in length
* 6315 6414: gap of 100 bp
* 6415 11824: contig of 5410 bp in length
* 11825 11924: gap of 100 bp
* 11925 20145: contig of 8221 bp in length
* 20146 20245: gap of 100 bp
* 20246 30884: contig of 10639 bp in length
* 30885 30984: gap of 100 bp
* 30985 44984: contig of 14000 bp in length
* 44985 45084: gap of 100 bp
* 45085 77877: contig of 32793 bp in length
* 77878 77977: gap of 100 bp
* 77978 112578: contig of 34601 bp in length
* 112579 112679: gap of 100 bp
* 112679 147259: contig of 34581 bp in length.
Location/Qualifiers
1. 147259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Rp11-115624"
/clone_lib="RPC1-II Human Male BAC"
1. 2950
/note="assembly_fragment"
3051. 6314
/note="assembly_fragment"
6415. 11824
/note="assembly_fragment"
11925. 20145
/note="assembly_fragment"
20246. 30884
/note="assembly_fragment"
30985. 44984
/note="assembly_fragment"
clone_end:17
vector_side:right
45085. 77877
/note="assembly_fragment"
77978. 112578

us-09-617-174b-1_1.rge

Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 RP11-416D8 is from the library RPCI-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-416D8 The true
 left end of clone RP11-320B8 is at 101956 in this sequence.

FEATURES

Source

```

1..184703
/organism="Homo sapiens"
/dbx_xref="taxon:9606"
/chromosome="10"
/clone="RP11-416D8"
/clone_lib="RPCI-11.2"

repeat_region
781..1138
/note="MLT1A2 repeat: matches 3..374 of consensus"
repeat_region
1378..1520
/note="MIR repeat: matches 13..169 of consensus"
repeat_region
1700..2027
/note="AluX repeat: matches 1..311 of consensus"
repeat_region
3259..3643
/note="MLT1B repeat: matches 1..390 of consensus"
repeat_region
4028..4397
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region
4934..5497
/note="L1ME repeat: matches 866..1405 of consensus"
repeat_region
5955..6285
/note="L1MEC repeat: matches 1813..2148 of consensus"
repeat_region
6335..6655
/note="L1M4 repeat: matches 2690..3056 of consensus"
repeat_region
6656..6959
/note="AluY repeat: matches 1..299 of consensus"
repeat_region
6960..7059
/note="L1M4 repeat: matches 3056..3146 of consensus"
repeat_region
7199..7937
/note="L1PAB repeat: matches 5420..6156 of consensus"
repeat_region
7950..8103
/note="AluY repeat: matches 158..311 of consensus"
repeat_region
8106..8401
/note="AluSp repeat: matches 1..304 of consensus"
repeat_region
8420..8744
/note="L1M4 repeat: matches 3412..3788 of consensus"
repeat_region
8760..9851
/note="L1M4 repeat: matches 3904..5041 of consensus"
repeat_region
9847..10012
/note="L1MB7 repeat: matches 5979..6149 of consensus"
repeat_region
10161..10335
/note="L2 repeat: matches 2019..2198 of consensus"
repeat_region
11843..11895
/note="MIR repeat: matches 85..145 of consensus"
repeat_region
12341..12649
/note="AluJo repeat: matches 1..309 of consensus"
repeat_region
13711..13861
/note="MIR repeat: matches 99..257 of consensus"
repeat_region
13868..14232
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region
14240..14328
/note="MIR repeat: matches 1..92 of consensus"
repeat_region
15157..15178
/note="11 copies 2 mer tt 100% conserved"
repeat_region
15659..16253
/note="L2 repeat: matches 1536..2155 of consensus"
repeat_region
16254..16535
/note="AluJo repeat: matches 13..293 of consensus"
repeat_region
16536..16584
/note="L2 repeat: matches 2155..2201 of consensus"
repeat_region
16585..17051
/note="MLT1D repeat: matches 24..505 of consensus"
repeat_region
17056..17346
/note="L2 repeat: matches 2196..2489 of consensus"
repeat_region
17471..17519
/note="L2 repeat: matches 2660..2708 of consensus"

```

```

repeat_region
18417..18641
/note="MIR repeat: matches 14..262 of consensus"
repeat_region
19232..19341
/note="MER5B repeat: matches 1..109 of consensus"
repeat_region
19355..19657
/note="AluJb repeat: matches 1..303 of consensus"
repeat_region
19935..20010
/note="MER3 repeat: matches 5..82 of consensus"
repeat_region
20104..20153
/note="L2 repeat: matches 2627..2676 of consensus"
repeat_region
22267..22495
/note="AluJo repeat: matches 57..293 of consensus"
repeat_region
23607..23909
/note="AluJb repeat: matches 5..306 of consensus"
repeat_region
23950..24341
/note="L1MAB repeat: matches 5899..6280 of consensus"
repeat_region
24645..24839
/note="L1MB2 repeat: matches 5962..6164 of consensus"
repeat_region
24853..25290
/note="Tigger2a repeat: matches 1..434 of consensus"
repeat_region
25291..25478
/note="94 copies 2 mer tg 61% conserved"
repeat_region
25485..25552
/note="34 copies 2 mer gt 66% conserved"
repeat_region
25491..25610
/note="3 copies 40 mer 79% conserved"
repeat_region
26969..27179
/note="MIR repeat: matches 20..232 of consensus"
repeat_region
27709..28003
/note="AluJb repeat: matches 1..288 of consensus"
repeat_region
28719..29028
/note="AluSg repeat: matches 1..311 of consensus"
repeat_region
29426..29730
/note="AluX repeat: matches 4..307 of consensus"
repeat_region
29738..29855
/note="L2 repeat: matches 2546..2654 of consensus"
repeat_region
30066..30366
/note="AluY repeat: matches 1..301 of consensus"
repeat_region
30370..30691
/note="L2 repeat: matches 2381..2709 of consensus"
repeat_region
31249..31513
/note="AluSc repeat: matches 25..306 of consensus"
repeat_region
31709..32120
/note="NSTA repeat: matches 1..424 of consensus"
repeat_region
33825..33856
/note="16 copies 2 mer ac 87% conserved"
repeat_region
34925..35218
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region
35223..35538
/note="AluSg repeat: matches 1..309 of consensus"
repeat_region
37794..37990
/note="MER3 repeat: matches 2..190 of consensus"
repeat_region
38000..38172
/note="MER5A repeat: matches 11..189 of consensus"
repeat_region
39449..39470
/note="11 copies 2 mer aa 100% conserved"
repeat_region
41110..41165
/note="28 copies 2 mer aa 75% conserved"
repeat_region
41416..41715
/note="AluJb repeat: matches 1..308 of consensus"
repeat_region
42825..43079
/note="AluX repeat: matches 44..301 of consensus"
repeat_region
43309..43558
/note="AluJ/FLAM repeat: matches 1..310 of consensus"
repeat_region
45407..45480
/note="Alu repeat: matches 241..301 of consensus"
repeat_region
46510..46702
/note="AluJo repeat: matches 86..302 of consensus"
repeat_region
48748..49033
/note="AluJo repeat: matches 1..301 of consensus"
repeat_region
50196..50327
/note="FLAM_C repeat: matches 1..132 of consensus"
repeat_region
51396..51701

```



```

O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Traversan,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zalcuncun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 21, 2001 this sequence version replaced gi:17149776.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- GenBank -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L9543
Center clone name: 69_M_11
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 154515: contig of 154515 bp in length
* 154516 154615: gap of 100 bp
* 154616 169669: contig of 15054 bp in length.
FEATURES             Location/Qualifiers
     source            1..169669
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-69M11"
BASE COUNT   46464 a 35508 c 36509 g 51088 t    100 others
ORIGIN
Query Match      17.0%; Score 194; DB 2; Length 169669;
Best Local Similarity 77.1%; Pred. No. 1.6e-43;
Matches 236; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 146 agagctgcgtactccttttcttttcttctgcttcacccaggctgaagtaca 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156615 AGAGTATGTTTTCTTTTGTGATACATGATCTTGTCCTGGCCAGGCTGGAGTGA 156556

Qy 206 gtggttagctcacggctcactcgagcttttgacctcccaggctcaagtatctctcgtct 265
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156555 GTGGTCAATCTCGGCTCCTACTGCACAACTTGGCTCTTTGGGCTCAAGTATTCTCTCACCT 156496

Qy 266 cagctttccaagttaactgagaccacaggcatgatcaccaacctattgatttttttaca 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156495 CAGCCTCCCAGTAGCTGGTAATTTACAAGCACATCACCACACCTGGGTATGTTTGTGA 156436

Qy 326 ttttttgtagagatgggtgtctcacatgttgcaccaggttggtctcaaaactctctgggtctca 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156435 TTTTGTAGAGAAGGGGTCTCTACTACATTTGCCAGACTGGTCTCAAACCTCTCGGACTCA 156376

Qy 386 agcaatccgtcacgtcacaacctccccaaaatgctgggattacagcgctgagccacccgggcc 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156375 AGCAATCTGCCACCTCAGCCTCCCCAAAAGTGTGGGATCACAGGTGTGAGCCACCACCATGCC 156316

Qy 446 agggct 451
|||||

```

```

/note="assembly_fragment"
misc_feature          95394..119296
/note="assembly_fragment"
misc_feature          119397..147115
/note="assembly_fragment"
misc_feature          147216..164126
/note="assembly_fragment"
vector_side:right"
clone_end:77
BASE COUNT   45794 a 34572 c 34173 g 48987 t    600 others
ORIGIN
Query Match      17.0%; Score 194; DB 2; Length 164126;
Best Local Similarity 77.1%; Pred. NO. 1.6e-43;
Matches 236; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

Qy 146 agagctgcgtactccttttcttttcttctgcttcacccaggctgaagtaca 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120866 AGAGTATGTTTTCTTTTGTGATACATGATCTTGCTGTGTCAGGCTGGAGTGA 120807

Qy 206 gtggttagctcacggctcactcgagcttttgacctcccaggctcaagtatctctcgtct 265
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120806 GTGGTCAATCTCGGCTCCTACTGCACATTTGGCTCTTGGGCTCAAGTATTCTCACCT 120747

Qy 266 cagctttccaagttaactgagaccacaggcatgatcaccaacctattgattttttaca 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120746 CAGCCTCCCAGTAGCTGGTAATTTACAAGCACATCACCACACCTGGCTAATGTTTGTGA 120687

Qy 326 ttttttgtagagatgggtgtctcacatgttgcaccaggttggtctcaaaactctgggtctca 385
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120686 TTTTGTAGAGAAGGGGTCTCTACTACATTTGCCAGACTGGTCTCAAACCTCTCGGACTCA 120627

Qy 386 agcaatccgtcacgtcacaacctccccaaaatgctgggattacagcgctgagccacccgggcc 445
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 120626 AGCAATCTGCCACCTCAGCCTCCCCAAAAGTGTGGGATCACAGGTGTGAGCCACCACCATGCC 120567

Qy 446 agggct 451
|||||
Db 120566 CAGCCT 120561

```

```

RESULT 15
AC041005/c
LOCUS
DEFINITION
Homo sapiens chromosome 8 clone RP11-69M11 map 8, HTG 07-FEB-2002
PROGRESS ***, 2 ordered pieces.
ACCESSION AC041005
VERSION AC041005.10 GI:17975388
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169669)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,I., Bouckhalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,K.,
Melrim,J., Maneus,L., Mihova,T., Miranda,C., Mlenga.V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

```

Db 156315 CAGCCT 156310

Search completed: September 24, 2002, 15:48:01
Job time: 19976 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 11:33:25 ; Search time 403.26 Seconds
(without alignments)
4857.907 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagagagag.....aggataactgtactccagg 1141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1137.8	99.7	1141	Human maspin promo
2	745.4	65.3	12426	Tumour suppressor
3	723.4	63.4	12426	Tumour suppressor
4	195	17.1	568	Human cDNA clone (
5	195	17.1	2343	Human cDNA sequenc
6	185.4	16.2	916	Human reproductive
7	185.4	16.2	32184	Human nervous syst
8	185.4	16.2	32184	Human nervous syst
9	185.4	16.2	36272	Human immune/haema

c 11	185.4	16.2	38348	22	AAK84953	Human immune/haema
c 12	184.6	16.2	240825	21	AAF24497	Human PG-3 gene.
c 13	184	16.1	1162	21	AAF79682	Human secreted pro
c 14	183.4	16.1	735	22	AAK05557	Human reproductive
c 15	183.4	16.1	735	22	AAK09703	Human digestive sy
c 16	182.6	16.0	1956	22	ABA21277	Human nervous syst
c 17	182.2	16.0	17026	22	AAK70372	Human immune/haema
c 18	181.2	15.9	1519	20	AA210674	cDNA encoding a hu
c 19	180.8	15.8	1421	22	AAK73297	Human immune/haema
c 20	180.4	15.8	21477	22	AAK66626	Human immune/haema
c 21	179.8	15.8	21480	22	AAK66625	Human immune/haema
c 22	179.8	15.8	1423	22	AAF27785	Human MANGO 347 co
c 23	179.8	15.8	1440	21	AAC90029	Clone HTHCW70 codi
c 24	179.2	15.8	1442	22	AAH33267	Human colon cancer
c 25	179.2	15.7	5269	22	AAK80329	Human immune/haema
c 26	178.8	15.7	7233	22	AAK36623	Human cardiovascular
c 27	178.8	15.7	8053	22	AAI98931	Human excretory re
c 28	178.8	15.7	8053	22	AAI63281	Human kidney relat
c 29	178.8	15.7	10241	22	AAK83768	Human immune/haema
c 30	178.6	15.7	1125	22	AAS33394	DNA encoding human
c 31	178.6	15.7	1125	22	ABA06543	Human cDNA SEQ ID
c 32	178.4	15.6	16181	22	AAK28831	Human immunoglobul
c 33	178	15.6	5962	23	AAK04268	Human reproductive
c 34	178	15.6	5962	23	AAS82348	DNA encoding novel
c 35	177.8	15.6	700	20	AAX30339	DNA encoding a hum
c 36	177.6	15.6	4071	22	AAS36540	Human cardiovascular
c 37	177.6	15.6	4071	22	AAK85200	Human immune/haema
c 38	177.2	15.5	21458	22	AAK06122	Human reproductive
c 39	177.2	15.5	21458	22	AAK81266	Human immune/haema
c 40	177	15.5	596	22	AAK75912	Human immune/haema
c 41	177	15.5	596	22	AAK75914	Human immune/haema
c 42	176.8	15.5	4431	22	ABA18399	Human nervous syst
c 43	176.8	15.5	4431	22	AAS40513	DNA encoding human
c 44	176.8	15.5	4431	22	AAK04139	Human reproductive
c 45	176.8	15.5	4433	22	ABA18400	Human nervous syst

ALIGNMENTS

RESULT 1
AAT89000
ID AAT89000 standard; cDNA: 1141 BP.

XX AC AAT89000;

XX DT 22-APR-1998 (first entry)

XX DE Human maspin promoter and partial coding cDNA.

XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;
tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT promoter 1..1141 ;

FT FT /*tag= a
/*note= "maspin promoter region"

FT FT misc_signal 443..449
/*tag= b

FT FT misc_signal 451..457
/*tag= c

FT FT misc_signal 485..490
/*tag= d

FT FT misc_signal 684..691
/*tag= e

FT FT misc_signal 846..851
/*tag= e

FT /tag= f
FT /note= "Putative regulatory element Ets"
FT misc_signal 907..913
FT /tag= g
FT /note= "Putative regulatory element AP1"
FT prim_transcript 958..1141
FT /tag= h
FT /note= "partial"
XX
XX
PN WO9736179-A1.
XX
XX 02-OCT-1997.
XX
XX 28-MAR-1997; 97WO-US05186.
XX
XX 28-MAR-1996; 96US-0014368.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX (DAND) DANA FARBER CANCER INST.
XX (PARD/) PARDEE A.
XX
XX Sagar R, Zhang M;
XX WPI; 1997-489785/45.
XX
XX Maspin gene promoter fragment - used to identify compounds for
XX treatment of prostate or breast cancer
XX
XX Claim 4; Fig 3; 51pp; English.

CC This sequence encodes the human maspin promoter region, including the
CC transcription start site. Maspin is a serpin which is expressed in
CC mammary epithelial cells. Its expression in these cells decreases
CC with increasing malignancy and is lost in during metastasis. Maspin
CC protein is also known to inhibit the mobility of tumour cells. This gene
CC can be used in method for screening compounds to identify candidate
CC compounds for the treatment of screening compounds to identify candidate
CC compounds for the treatment of prostate cancer, or breast cancer. It can
CC also be used to identify compounds that increase the expression of
CC maspin, and for detecting the presence of metastatic prostate epithelial
CC cells.

XX Sequence 1141 BP; 261 A; 292 C; 277 G; 311 T; 0 other;

Query Match 99.7%; Score 1137.8; DB 18; Length 1141;
Best Local Similarity 99.8%; Pred. No. 2.4e-301;
Matches 1139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 agataagcacagagagaagaacacagctcgtttcaggtcctttcctgaggtgattcg 60
Db 1 agataagcacagagagaagaacacagctcgtttcaggtcctttcctgaggtgattcg 60
Qy 61 gctggaaggagtaggttccaccacaaatgaagaagctgtggaagacagaggagacaagaac 120
Db 61 gctggaaggagtaggttccaccacaaatgaagaagctgtggaagacagaggagacaagaac 120
Qy 121 aggtctccacgaagagatttcagagacagctgcgtactccttttctttttttttttt 180
Db 121 aggtctccacgaagagatttcagagacagctgcgtactccttttctttttttttttt 180
Qy 181 gctctgtcaccacagctgaagtagacagtggttagctcagggctcactgcagctttgacctc 240
Db 181 gctctgtcaccacagctgaagtagacagtggttagctcagggctcactgcagctttgacctc 240
Qy 241 ccaggctcaagtatctctcgtctcagctttccaaagtaactgggacacagagcatgcat 300
Db 241 ccaggctcaagtatctctcgtctcagctttccaaagtaactgggacacagagcatgcat 300
Qy 301 caccacgctaggtattgttttacattttttgtagagatggggtctcaccatgttgccca 360
Db 301 caccacgctaggtattgttttacattttttgtagagatggggtctcaccatgttgccca 360
Qy 361 ggttggtctcaaacctcgtggtcgaagcaatccgctcagctcaacccctcccccaatgctgg 420

Db 361 ggttggtctcaaacctcgtggtcgaagcaatccgctcagctcgaacccctcccccaatgctgg 420
Qy 421 gattacagggctgagccaccggccagggctgagtaactccttaatacacaggattttaaaaa 480
Db 421 gattacagggctgagccaccggccagggctgagtaactccttaatacacaggattttaaaaa 480
Qy 481 gaaacttctgcgccaccatataacataatctctacaaatttggtagtaaatatttg 540
Db 481 gaaacttctgcgccaccatataacataatctctacaaatttggtagtaaatatttg 540
Qy 541 ctaatagtagtaacttaatttttagtaggacctgtgtttatatacatatcatctctcttt 600
Db 541 ctaatagtagtaacttaatttttagtaggacctgtgtttatatacatatcatctctcttt 600
Qy 601 ttgattgtcttctgttttaattgggagctacacctctcttggtcgtctagcagaatgagctgc 660
Db 601 ttgattgtcttctgttttaattgggagctacacctctcttggtcgtctagcagaatgagctgc 660
Qy 661 tgcagtttacacaaaaaagaatggagatcagagtagcttttttgcaccacaactgtgtgag 720
Db 661 tgcagtttacacaaaaaagaatggagatcagagtagcttttttgcaccacaactgtgtgag 720
Qy 721 aaattgttagttactatcatcacacattacttttttatttctcgaataatttcaactcc 780
Db 721 aaattgttagttactatcatcacacattacttttttatttctcgaataatttcaactcc 780
Qy 781 ggtcctgcgtggggccgagagagattgccgtacgcatgtctgtacgtatgcatgtaactcac 840
Db 781 ggtcctgcgtggggccgagagagattgccgtacgcatgtctgtacgtatgcatgtaactcac 840
Qy 841 agccctctcgtcccgaaacattgttgagagccttttgaagctgtgcagacaacagcaact 900
Db 841 agccctctcgtcccgaaacattgttgagagccttttgaagctgtgcagacaacagcaact 900
Qy 901 tcagctgaatcatctcttcaattgtgacaagctgccaagaggtttagtaggagaggg 960
Db 901 tcagctgaatcatctcttcaattgtgacaagctgccaagaggtttagtaggagaggg 960
Qy 961 agtgccgcgaggg 1020
Db 961 agtgccgcgaggg 1020
Qy 1021 gctgccaggtgaagccaccgctgtcttctccacagacacggtcgtccctccacatccaggtct 1080
Db 1021 gctgccaggtgaagccaccgctgtcttctccacagacacggtcgtccctccacatccaggtct 1080
Qy 1081 ttgtgctcctcgtctgctgttctcttccacgcatctttccaggataactgtgactccag 1140
Db 1081 ttgtgctcctcgtctgctgttctcttccacgcatctttccaggataactgtgactccag 1140
Qy 1141 g 1141
Db 1141 g 1141
RESULT 2
AAS46482/C
ID AAS46482 standard; DNA; 12426 BP.
XX
XX AAS46482;
XX
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #204.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
PN

XX PD 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-EP02955.

XX PR 15-MAR-2000; 2000DE-1013847.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2001-602752/68.

XX PT Fragments of chemically modified genes associated with tumour suppressor

XX PT genes and oncogenes, useful in designing primers and probes for

XX PT analysing diseases associated with cytosine methylation state e.g.

XX PT cancer -

XX PS Claim 1; SEQ ID No 204; 27pp; English.

XX CC The invention relates to a nucleic acid comprising a sequence of 18

XX CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

XX CC bisulphite, of genes associated with tumour suppression and

XX CC oncogenes having a sequence taken from 536 (actually 533 since

XX CC numbers 408, 458 and 500 are missing from the sequence listing) sequences

XX CC (SS) and sequences complementary to (SS). The nucleic acid may be a

XX CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

XX CC form part of a set of probes for detecting the cytosine methylation state

XX CC and/or single nucleotide polymorphisms and also to be used in an

XX CC array for analysing diseases associated with CpG dinucleotides e.g.

XX CC cancers and tumours. The probes can also be used in a method for

XX CC ascertaining genetic and/or epigenetic parameters for the diagnosis

XX CC and/or therapy of existing diseases or the predisposition to specific

XX CC diseases, by analysing cytosine methylations. The parameters may be

XX CC compared to another set of genetic and/or epigenetic parameters, the

XX CC differences serving as basis for diagnosis and/or prognosis events which

XX CC are disadvantageous to patients. The present sequence is one of the

XX CC 533 genomic sequences derived from tumour suppressor genes and

XX CC oncogenes. Sequences with even numbered Seq ID numbers are the

XX CC complementary sequence of the corresponding odd numbered sequence (e.g.

XX CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence

XX CC is missing).

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

Query Match 65.3%; Score 745.4; DB 22; Length 12426;

Best Local Similarity 78.4%; Pred. No. 1.9e-193;

Matches 893; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

QY 1 agataagcacagcagagaagcaaccagctccgtttcaggtctcttctcaggtcgtattcg 60

DB 8506 AATAAACACACAAAAAACAACAACTCCCGTTTCAAAATCCTTTCTAAAACTAATTCG 8447

QY 61 gctggaggaggtagtctccaccacaaatgagaagctgtggaagcagagagagacaagac 120

DB 8446 ACTAAAAAATAATCCCGCAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 8387

QY 121 aggtccacagagagatccagagcagagctgcgtactcttcttcttcttcttcttcttctt 180

DB 8386 AAACTCCACGAAAAAATTCAAAAACAATACTAGTACTCCTTTCTTTTATTCTTTT 8327

QY 181 gctctgtcaccagcagctgaagtacagtggttagctcacgctcagctcagcttgacctc 240

DB 8326 ACTCTATCACGCAAACTAAAAATACAATAATTAACACCACTCACTACAACTTTAACCTC 8267

QY 241 ccaggctcaagtgatcctctcgtctcagcttttccaaagtaactgaggaccacagcagcatgc 300

DB 8266 CCAAACTCAAAATAATCTCTCGTCTCAACTTTCCAAATAACTAAACACCAACATACAT 8207

QY 301 caccacgctagcgtatctgtttttacattttttgtagagatgagggctcaccacatgtctccca 360

DB 8206 CACCACATAAACTATATTTTATACATTTTATATAAAATAAAATCTCACCATATTAACCA 8147

QY 361 ggttggtctcaaaactccttggtcctcaagcaatccgctcagctcaacccctcccaaatgctgg 420

DB 8146 AATTAACTCAAACTCTTAACTCAAACTCCGCTCAGCTCAACCTCCCAATAACTAA 8087

QY 421 gattacagcgtgagccacggggcagggtgagtaactcctaatacagagatttttaaaaa 480

DB 8086 AATTCAAAACGTAACACCGCCGCAAACTAAATAATCTTAATCAAAAAATTTAAAAA 8027

QY 481 gaaactcctgcgcacccaccattaaacaatatctcctaccacatttggttagtaataatttg 540

DB 8026 AAAAATCTTCCACGCCACCCATTAACAATATCTCTTACCATAATTTAATAATAATTTTA 7967

QY 541 ctaaatagctacaaatttttaggtagcactgtgtttatcacatatatccattcctctttt 600

DB 7966 CTAAATAATACCTAAATTTTAAATAAACAATATATTTATACATATATCCATTCCTCTTT 7907

QY 601 ttgattgtcttctgttttaagtggcagctacctctcttggcatctgacagaaatgagctgc 660

DB 7906 TTAATTTATCTTCTATTTAATAAACAATACCTCTCTTAACATCTAACAATAAATACTAC 7847

QY 661 tgcagtttacacaaaagaatggagatcagagtagtactttttgtgccaccacagtgctgag 720

DB 7846 TACAATTTTACACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7787

QY 721 aaatttgtagtgttactatcatcacattacttttttttttttttttttttttttttttttt 780

DB 7786 AAATTTATATAATATCTATCATCATCACATTTACTTTTATTTTCATCGAATATTTTCACTTC 7727

QY 781 ggtcctcgtcggggcagagagattggtgacgcatgctgacgtatgcatgtgcatgtaactcac 840

DB 7726 GATCTCTACGTAAACCCGCAAAAAATACCGTCAGCATATCTATACGTATACATATACTAC 7667

QY 841 agcccttctcgtccgcaacatgttggaagcctttttggaagctgtgagacacacagaact 900

DB 7666 AACCCCTCTCTACCCGCAACATATTAATAAATCTTTTAAAAAATCTATACAAACAATAA 7607

QY 901 tcagcctgaatcatctcttcaattgtgacaagctgccaagaggttgagtaggagag 960

DB 7606 TCAACCTAAATCATTTCTTCAATTTAACAACAACTACCAAAAAAATTAATAAAAAA 7547

QY 961 agtgcgcgcagagcggggcggggcggggcggggcggggcggggcggggcggggcgggg 1020

DB 7546 AATACCGCGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 7487

QY 1021 gctgcccaggtgagccacccgctgtcttctgcccagacagcggctccaccatccaggtct 1080

DB 7486 ACTACCAATAAACCACCGCTACTTCTTACCCAAACAGCATCGCTCCACATCAATACT 7427

QY 1081 ttgtgctcctcgtctgctgtctcttctccagcatttccagagataaactgtgaactcca 1139

DB 7426 TTATACTCTCTCGTTCCTTATCTTCTTTTCCACGCAATTTTCCAAAAATAAATAAATA 7368

RESULT 3'

AAS46481

ID AAS46481 standard; DNA: 12426 BP.

XX

AC AAS46481;

XX

DT 18-DEC-2001 (first entry)

XX

DE Tumour suppressor gene derived chemically modified sequence #203.

XX

KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 2343 BP; 603 A; 515 C; 535 G; 690 T; 0 other;

Query Match 17.1%; Score 195; DB 22; Length 2343;
Best Local Similarity 80.6%; Pred. No. 4.le-43;
Matches 228; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 177 ttttctctctaccagctgaagtacagtggttagctcaagctcaactgcagctttga 236
Db 2331 TCTTGTCTCTACCCAGCAGGAGTGCACTGCAGATCACTGCAGCTTTGA 2272
Qy 237 cctcccaggctcaagtgcctctctcagctttcccaagtaactgggaccacaggaat 296
Db 2271 CCTCCAGGCTCAGTGCATCTCCACCTCAGCTCCGAGTAGCTGGGACTACAGGCAT 2212
Qy 297 gcataccagctaggtattgtttacatttttttagagatggggtctcaccatgttg 356
Db 2211 GCACCACACGCTCCAGCTAATTTTGTACTTTTGTAGAGACGGGTTTCACCAITGTTG 2152
Qy 357 cccaggttggtctcaaacctctgggctcagcaatccgctcagctcaaacctccccaaatg 416
Db 2151 CACAGGCTGGTCTCAGACTCTGGGCTCAAGCTATCGGCTGCTTCGGCCTCCCAAAGTG 2092
Qy 417 ctgggattacagggcgtgagcaccggccagggcgtgagtaatc 459
Db 2091 CTGGGATTACAACTGGGAGCCACGCTGCGCCGCCGAGATCTC 2049

RESULT 6
AAL05530/c
ID AAL05530 standard; DNA; 916 BP.
XX
XX AAL05530;
XX
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8218.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225269.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-541565/60.
Nucleic acids encoding 3224 human nervous system antigen polypeptides,
useful for preventing, diagnosing and/or treating nervous system
cancers and metastases -
Disclosure; SEQ ID NO 12920; 1701pp + Sequence Listing; English.
The invention relates to novel genes (ABA11004-ABA21534) and proteins
(ABBI4678-ABBI8001) useful for preventing, treating or ameliorating
medical conditions e.g. by protein or gene therapy. The genes are
isolated from a range of human tissues disclosed in the specification.
The nucleic acids, proteins, antibodies and (ant)agonists are useful
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
(f) immune disorders e.g. Addison's disease, allergies, autoimmune
haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
disease, multiple sclerosis, rheumatoid arthritis and ulcerative
colitis; (c) cardiovascular disorders such as myocardial ischaemias;
(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 32184 BP; 7866 A; 6152 C; 6798 G; 11368 T; 0 other;

Query Match 16.2%; Score 185.4; DB 22; Length 32184;
Best Local Similarity 74.3%; Pred. No. 5.8e-40;

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0230160.
PR 01-DEC-2000; 2000US-0230391.
PR 05-DEC-2000; 2000US-0231030.
PR 05-DEC-2000; 2000US-0231988.
PR 05-DEC-2000; 2000US-02356719.
PR 06-DEC-2000; 2000US-0231479.
PR 08-DEC-2000; 2000US-0231856.
PR 08-DEC-2000; 2000US-0231868.
PR 08-DEC-2000; 2000US-0231869.
PR 08-DEC-2000; 2000US-0231989.
PR 08-DEC-2000; 2000US-0231990.
PR 11-DEC-2000; 2000US-0234097.
PR 05-JAN-2001; 2001US-02359678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure: SEQ ID NO 39764; 3071pp + Sequence Listing; English.
PS
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients' own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic


```
FT primer_bind /tag= i 4559..4577
FT primer_bind /tag= j 4582..4600
FT misc_binding /tag= k 4589..4613
FT allele /tag= l /note= "binds probe"
FT primer_bind replace(4601,G)
FT primer_bind /tag= m 40242..40259
FT exon /tag= n 4627..4718
FT primer_bind /tag= o 4891..4908
FT primer_bind /tag= p 10007..10025
FT exon /tag= q 10115..10233
FT primer_bind /tag= r /label= "B"
FT primer_bind /tag= s 10209..10227
FT misc_binding /tag= t 10216..10240
FT allele /note= "binds probe"
FT primer_bind replace(10228,T)
FT primer_bind /tag= complement(10229..10247)
FT primer_bind /tag= v 10267..10285
FT misc_binding /tag= w 10274..10298
FT allele /tag= x /note= "binds probe"
FT primer_bind replace(10286,T)
FT allele /tag= y /note= "binds probe"
FT primer_bind complement(10287..10305)
FT allele /tag= z /tag= complement(10370,)
FT primer_bind 10411..10430
FT exon /tag= ab 26810..26897
FT exon /tag= ac /label= "D"
FT exon 31357..31471
FT exon /tag= ad /label= "E"
FT exon 34261..34404
FT exon /tag= ae /label= "F"
FT exon 37377..37466
FT exon /tag= af /label= "S"
FT exon 39704..40858
FT primer_bind /tag= ag /label= "T"
FT primer_bind 39556..39574
FT primer_bind /tag= ah 39877..39896
FT primer_bind /tag= ai 39925..39943
FT misc_binding /tag= aj 39932..39956
FT allele /tag= ak /note= "binds probe"
FT primer_bind replace(39944,T)
FT primer_bind /tag= al complement(39945..39963)
FT primer_bind /tag= am 39953..39970
FT primer_bind /tag= an 39954..39972
FT misc_binding /tag= ao 39961..39985
FT allele /tag= ap /note= "binds probe"
FT primer_bind replace(39973,C)
FT primer_bind /tag= aq complement(39974..39992)
FT primer_bind /tag= ar 40242..40259
FT primer_bind /tag= as 41137..41154
FT primer_bind /tag= at 41366..41384
FT misc_binding /tag= au 41373..71397
FT allele /tag= av /note= "binds probe"
FT allele replace(41385,C)
FT primer_bind /tag= aw 41385..41403
FT primer_bind /tag= ax 41385..41403
FT misc_binding complement(41386..41404)
FT allele /tag= ay 41392..41416
FT primer_bind /tag= az /note= "binds probe"
FT primer_bind replace(41404,C)
FT primer_bind /tag= ba complement(41405..41423)
FT primer_bind /tag= bb 41564..41581
FT primer_bind /tag= bc 42122..42141
FT primer_bind /tag= bd 42213..42231
FT misc_binding /tag= be 42220..42244
FT allele /tag= bf /note= "binds probe"
FT primer_bind replace(42232,C)
FT primer_bind /tag= bg complement(42233..42251)
FT primer_bind /tag= bh 42526..42543
FT exon /tag= bi 50436..50545
FT primer_bind /tag= bj /label= "G"
FT primer_bind 67289..67309
FT primer_bind /tag= bk 67456..67474
FT misc_binding /tag= bl 67463..67487
FT allele /tag= bm /note= "binds probe"
FT primer_bind replace(67475,G)
FT primer_bind /tag= bn complement(67476..67494)
FT primer_bind /tag= bo 67724..67741
FT primer_bind /tag= bp 69182..69200
FT primer_bind /tag= bq 69502..69520
FT misc_binding /tag= br 69509..69533
FT allele /tag= bs /note= "binds probe"
FT primer_bind replace(69521,G)
FT primer_bind /tag= bt complement(69522..69540)
```


XX
AC AAL05557;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 8245.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO20015320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214986.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225288.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
PS Disclosure; SEQ ID NO 13608; 170lpp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AB11004-ABA21534) and proteins
CC (AB14678-ABB18001) useful for preventing, treating or ameliorating
CC

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 10:07:45 ; Search time 3309.28 seconds
(without alignments)
4653.588 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagagagaag.....aggataactgtgactccagg 1141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	197.8	17.3	496	12 B35967	B35967 HS-1031-A2-
c 2	195	17.1	568	9 AU159406	AU159406 AU159406
c 3	186.2	16.3	410	9 AI023375	AI023375 ow63902.x
c 4	184.2	16.1	708	12 AG143040	AG143040 Pan trogl
c 5	184.2	16.1	713	12 AQ312217	AQ312217 RPCI11-10
c 6	183.6	16.1	655	12 AG169306	AG169306 Pan trogl
c 7	183	16.0	643	9 AI114719	AI114719 HAI1318 Hu
c 8	183	16.0	2167	11 AF116633	AF116633 Homo sapi
c 9	182.4	16.0	390	10 R56162	R56162 yg91b06.r1
c 10	182.4	16.0	742	12 AQ343483	AQ343483 RPCI11-12
c 11	181.8	15.9	649	12 BH609847	BH609847 HIV13C04
c 12	181.6	15.9	524	9 AW973992	AW973992 EST385990
c 13	180.6	15.8	571	9 AA131088	AA131088 zo16c04.s
c 14	179.8	15.8	331	10 BG057233	BG057233 7175d01.x
c 15	179.8	15.8	661	9 BE177530	BE177530 RCI-HT059
c 16	179.8	15.8	698	12 AG141964	AG141964 Pan trogl
c 17	179.6	15.7	430	12 AQ393433	AQ393433 CITBI-E1-

18	179.4	15.7	424	9 AI499954	AI499954 to63c10.x
19	179.2	15.7	373	9 AL596605	AL596605 DKF2p451C
20	179.2	15.7	559	10 BG537830	BG537830 602566186
21	179.2	15.7	713	10 BG684208	BG684208 602635838
22	179.2	15.7	893	10 BI913559	BI913559 603180311
c 23	179.8	15.7	1048	9 AL576528	AL576528 AL576528
c 24	178.8	15.7	2205	11 AF318346	AF318346 Homo sapi
c 25	178	15.6	455	12 AQ392348	AQ392348 CITBI-E1-
c 26	178	15.6	558	12 AQ527359	AQ527359 RPCI-11-3
c 27	177.8	15.6	379	9 AV754799	AV754799 AV754799
c 28	177.8	15.6	684	12 AG037064	AG037064 Pan trogl
c 29	177.4	15.5	456	9 AI753334	AI753334 cr08f09.x
c 30	177.4	15.5	512	12 AQ538435	AQ538435 RPCI-11-3
c 31	177.2	15.5	416	12 AQ748789	AQ748789 HS_5574_A
c 32	177.2	15.5	515	12 AQ506282	AQ506282 RPCI-11-3
c 33	177	15.5	532	12 AQ701252	AQ701252 HS_2132_A
c 34	177	15.5	632	9 AI821267	AI821267 ab90f07.x
c 35	176.6	15.5	562	12 AQ778690	AQ778690 HS_2235_A
c 36	176.6	15.5	701	12 AG148726	AG148726 Pan trogl
c 37	176.4	15.5	512	12 AQ120174	AQ120174 HS_3028_B
c 38	176.4	15.5	950	12 AQ749313	AQ749313 HS_5576_A
c 39	176.2	15.4	471	9 AW151925	AW151925 xf70d02.x
c 40	176.2	15.4	731	12 AG175563	AG175563 Pan trogl
c 41	176.2	15.4	900	10 BE880692	BE880692 601490306
c 42	176.2	15.4	1020	10 BG615396	BG615396 602643974
c 43	176	15.4	387	9 AA736485	AA736485 nv92h02.r
c 44	176	15.4	792	9 AU136365	AU136365 AU136365
c 45	175.8	15.4	463	12 AQ033859	AQ033859 HS_2194_B

ALIGNMENTS

RESULT 1

B35967/c
LOCUS HS-1031-A2-F07-MR-V0001.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=14 Row=K, DNA sequence.
DEFINITION B35967
ACCESSION B35967.1 GI:25353336
VERSION B35967
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS Mahairas G.G., Zackrone, K.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 811 row: K column: 14
Class: BAC ends
High quality sequence stop: 496.
FEATURES
Location/Qualifiers
1..496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 811 Col=14 Row=K"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 155 a 85 c 135 g 120 t 1 others
ORIGIN


```

Query Match          16.0%; Score 183; DB 9; Length 643;
Best Local Similarity 77.4%; Pred. No. 9.2e-29;
Matches 222; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 177 ttctgctgtcacccaggctgaagtacagtggttagctcacggctcactgcagctttga 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TCTTGCTCTGTCTACCCAGGCTGGAGTGCAGTGTGACAAATCTTGGCTCATAATAGCCTCGA 159

QY 237 cctccaggctcaagtgatcctctctgctcagctttccaaagtaactgggaccacagggcat 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 CTTCCAGGCTCAGGTGATCTCCCACTCAGCTCCCAACTAGCTGGGACACACAGGTGC 219

QY 297 gcatcacacgctaggtctattttttacatttttttagatagatggggtctcaccatgttg 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 CCGCCACACAGGCTGGCTATTTTGTATTTGTGTAGATAGGGGTTTGGCATGTGTG 279

QY 357 ccaggttggtctcaaacctcctgggtcgaagcaatccgcctcagctcaaacctccccaatg 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 GCATGCTGCTCGAAGCTCTGGGCTCAAGCAATCCACCATCTCAGCTCCCAAGTG 339

QY 417 ctgggattacaggtgagccaccgggcccagggtcgtgagtaactctaa 463
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 CTGGGATTACAGGCTGGGCGCCACACACTCGGCCCAACAATTTCTTAA 386

RESULT 8
R56162 AF116633 2167 bp mRNA linear HTC 08-MAY-2001
LOCUS Homo sapiens PRO1318 mRNA, complete cds.
DEFINITION AF116633
ACCESSION AF116633
VERSION AF116633.1 GI:7959767
KEYWORDS HTC
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2167)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
Bi,J., Liu,M. and He,F.
TITLE Functional prediction of the coding sequences of 121 new genes
deduced by analysis of cDNA clones from human fetal liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2167)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
Bi,J., Liu,M. and He,F.
TITLE Direct Submission
Submitted (24-DEC-1998) Department of Experimental Hematology,
Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing
100850, P. R. China
FEATURES
Location/Qualifiers
1..2167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLB5032"
/tissue_type="liver"
/dev_stage="fetal"
/complement(85..365)
/rpt_family="Alu"
/complement(411..543)
/rpt_family="MER45"
436..689
/note="predicted protein of HQ1318"
/codon_start=1
/product="PRO1318"
/protein_id="AAF71056.1"
/db_xref="GI:7959768"
/translation="MPSVAQGPVPHLGSRSNAVFEFLVMFEQRPYVILHWAPQIT
WFLRRGVSHQSPKSPLEVLNRTAEFLKSSVGETVHHHTQ"
repeat_region /rpt_family="Alu"
repeat_region 1137..1304
BASE COUNT 643 a 420 c 408 g 696 t

```

```

Query Match          16.0%; Score 183; DB 11; Length 2167;
Best Local Similarity 77.4%; Pred. No. 9.7e-29;
Matches 222; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 177 ttctgctgtcacccaggctgaagtacagtggttagctcacggctcactgcagctttga 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 TCTTGCTCTGTCTACCCAGGCTGGAGTGCAGTGTGACAAATCTTGGCTCATAATAGCCTCGA 159

QY 237 cctccaggctcaagtgatcctctctgctcagctttccaaagtaactgggaccacagggcat 296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 CTTCCAGGCTCAGGTGATCTCCCACTCAGCTCCCAACTAGCTGGGACACACAGGTGC 219

QY 297 gcatcacacgctaggtctattttttacatttttttagatagatggggtctcaccatgttg 356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 CCGCCACACAGGCTGGCTATTTTGTATTTGTGTAGATAGGGGTTTGGCATGTGTG 279

QY 357 ccaggttggtctcaaacctcctgggtcgaagcaatccgcctcagctcaaacctccccaatg 416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 GCATGCTGCTCGAAGCTCTGGGCTCAAGCAATCCACCATCTCAGCTCCCAAGTG 339

QY 417 ctgggattacaggtgagccaccgggcccagggtcgtgagtaactctaa 463
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 CTGGGATTACAGGCTGGGCGCCACACACTCGGCCCAACAATTTCTTAA 386

RESULT 9
R56162 R56162 390 bp mRNA linear EST 23-MAY-1995
LOCUS Yg91b06.r1 Soares infant brain lN1B Homo sapiens cDNA clone
DEFINITION IMAGE:40776 5' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION R56162
VERSION R56162.1 GI:826268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1266
High quality sequence stops: 349 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1266 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 349.
FEATURES
Location/Qualifiers
1..390
/organism="Homo sapiens"
/db_xref="GDB:413317"
/db_xref="taxon:9606"
/clone="IMAGE:40776"
/clone_lib="Soares infant brain lN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site:1: Not
1; Site_2: Hind III; 1st strand cDNA was primed with a Not

```


Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCAGCAG 3' ~3' adaptor sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3'									
BASE COUNT	135 a	130 c	137 g	161 t	8 others				
ORIGIN									
Query Match	15.8%;	Score 180.6;	DB 9;	Length 571;					
Best Local Similarity	78.5%;	Pred. No. 3e-28;							
Matches 216;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;					
QY	177	ttttgtctgtcacccaggctgaagtacagtgttagctcacggctcaactgcagctgttga	236						
Db	34	TCATTACTTGTTCACCCAGGCTGGAATCAATGGCAAGATCATGCTCACTCGACGCTCGA	93						
QY	237	ccctccaggctcaagtgaatcctctcgtctcagctttccaagtaactcaggaccacaggcat	296						
Db	94	CCTCCCAGGCTCAAGTGATCCTCCCATCTCAGCCTCCCGCTCCCGAGTGGGACCAAGCAT	153						
QY	297	gcataccacagctaggctattgttttaccattttttagagatggggtctcaccatgttg	356						
Db	154	GTGCCACCACACCTGGCTAAATTTTGTATTTTGTAGAGACAGGGFTTGGCATGTTG	213						
QY	357	cccagggtgtctcaaacctcctgggtcacaagcaatccgctcgaogtcaacctccccaaatg	416						
Db	214	GCCAGGCTGGTCTTGAACCTCCAGGCTCAAGCAATTCGCGTCCGCTGCTCCACAGTG	273						
QY	417	ctgggattacaggctgagcacccggccaggct	451						
Db	274	CTGGGATTACAGGCATGAGTCACTTTTCCTGGCCCT	308						
RESULT 14									
LOCUS	BG057233								
DEFINITION	7f75d01.x1 Lupski_dorsal_root_ganglion Homo sapiens cDNA clone	331 bp	mRNA	linear	EST 25-JAN-2001				
IMAGE:3302809 3' similar to contains Alu repetitive element;; mRNA sequence.									
ACCESSION	BG057233								
VERSION	BG057233.1	GI:12522512							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
REFERENCE	1 (bases 1 to 331)								
AUTHORS	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.							
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index								
JOURNAL	Unpublished (1997)								
COMMENT	Contact: Robert Strausberg, Ph.D.								
	Email: cgapsb@email.nih.gov								
	Tissue Procurement: Dr. James Lupski								
	cDNA Library Preparation: Lupski Laboratory								
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)								
	Cloning by: Washington University Genome Sequencing Center								
	Cloning by: NCI-CGAP	clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov							
	Seq primer: -400P from Gibco								
	High quality sequence stop: 312.								
	Location/Qualifiers								
FEATURES	1..331								
source	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/clone="IMAGE:3302809"								
	/clone_lib="Lupski_dorsal_root_ganglion"								
	/sex="male"								
	/tissue_type="dorsal root ganglia"								
	/dev_stage="adult, 36 yr"								
	/lab_host="DH10B"								
	/note="Vector: pCMV-SPORT6 (Life Technologies); Site:1"								

NotI; Site_2: SalI; cDNA made by oligo-dT priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGCGTCGG-3' and 5'-GACTAGTTCTAGATCGGAGCGGCCGCCT(15)-3'. Size selected > 1 kb for average insert length 1.7 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."					
BASE COUNT	78 a	88 c	73 g	92 t	
ORIGIN					
Query Match	15.8%	Score 179.8;	DB 10;	Length 331;	
Best Local Similarity	76.7%;	Pred. No. 4.3e-28;			
Matches 220;	Conservative	0;	Mismatches 67;	Indels 0;	Gaps 0;
Qy	177	ttttgcctgtcaccaggctgaagtacagtggttagctcacggctcaactgacgaccttga	236		
Db	12	TCTCACCCTGTCTCACCCAGGCTGGATCGGTTGTCATGATCAGAGCTCACTGCACGCCTTA	71		
Qy	237	cctccaggctccaagtatcctctcgctcagctttccaagtaactgggaccacagcat	296		
Db	72	CCTTCCAGGCTCAAGTGATCTCTCACTTCAGCTCCCTTGTAGCTGGGACCACAGGTGC	131		
Qy	297	gcataccacgctaggctattgttttacatttttttagagatggggtctcaccatgtg	356		
Db	132	GTGCCACACACTTGGCTAAATTTTTAAATTTTTTTGTAGAGATGGGTCTTGCTATGTTG	191		
Qy	357	cccagggttgtctcaactcctgggtcaaagaatcgctcagctcaactccccaaatg	416		
Db	192	CCCAGGTGGTCTCAAAATTCCTGGGCTCAAGTGCTCTTTCGCCTCAGCCTCCCAAAGTG	251		
Qy	417	ctgggattacaggcgtgagccacgggccaggcgtgagtaaatccta	463		
Db	252	CTGGGACTACAGACATGAGCTACTGGCCCCAGCCTTAGAAAACITTA	298		
RESULT 15					
BE177530/c					
LOCUS	BE177530	661 bp	mRNA	linear	EST 22-JUN-2000
DEFINITION	R01-HT0597-170300-021-a03 HT0597 Homo sapiens cDNA, mRNA sequence.				
ACCESSION	BE177530				
VERSION	BE177530.1 GI:865682				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zagó,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brundstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.				
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags				
JOURNAL	Proc.Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)				
MEDLINE	20202663				
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-R01-HT0597-170300-021-a03&t3=2000-03-17&t4=1) Seq primer: puc 18 forward				

High quality sequence start: 13
High quality sequence stop: 660.
Location/Qualifiers
1. .661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0597"
/dev_stage="Adult"
/note="Organ: head-neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 190 a 145 c 154 g 172 t
ORIGIN

Query Match 15.8%; Score 179.8; DB 9; Length 661;
Best Local Similarity 75.6%; Pred. No. 4.4e-28;
Matches 223; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
Qy 183 tctgtcaccaggctgaagtacagtggttagctcagcgctcactgcagctttgacctccc 242
Db 520 TTTGTGGCCAGCTGGAGTGCATGCTGGCTTTGGCTACCGCAACCTCCGCTGCC 461
Qy 243 aggtcgaagtgcctctcgtctcagcttccaaagtaactgggaccacagggcatcatca 302
Db 460 GGGTTCAAGTGATTCCTCGCTCAGCCTCCCGAGTAAC TGGGACCCACAGGCGCCACCA 401
Qy 303 ccacgctagggtattgttttacattttttgttagagatggggtctcaccatgttgcccagg 362
Db 400 CCACGCCCAGCTAATTTTGTGTAATTCATAGAGATGGGTTTCACCATGTTGCCAGG 341
Qy 363 ttgggtctcaaaactcctggggctcaagcaatcccgctcacgtcaacctccccaaatgctggga 422
Db 340 CTGGTCTAAACTCTGGCCCTCAAGTGATCCACCACTTCGGCCCTCCCAAGTCTGGGA 281
Qy 423 ttacagggcgtgagccacggggccagggtgagtaatcctctaatcacaggattttaa 477
Db 280 TTACAGGCGTGAGCCGCCACGCCAGCCTAGATGCAAGTTATTTTAAAGTGTA 226

Search completed: September 24, 2002, 14:45:10
Job time: 16645 sec

THIS PAGE BLANK

THIS PAGE BLANK (USPTO)